School of Basic and Applied Sciences

Department of Mathematical Sciences

Programme Name: B.Sc (Applied Statistics and Data Science)

Programme Code: MTH3403

Undergraduate Course Structure

Batch: 2026-2030

SCHOOL OF BASIC AND APPLIED SCIENCES
UNDERGRADUATE COURSE STRUCTURE
B.Sc (Applied Statistics and Data Science)

SEMESTER I								
SL. No	**Type of Course**	**Code**	**Title of the Course**	**Credit Distribution**				**Remarks**
				L	**T**	**P**	**C**	
1	CC	SDS101	Probability Theory	3	1	0	4	CC-1
2	CC	SDS102	Descriptive Statistics	3	0	1	4	CC-2
3	Minor	IKS101	Introduction to IKS	4	0	0	4	
4	MDC						3	
5	AEC	AEC101	Communicative English I	2	0	0	2	
6	AEC		Foreign Language I	2	0	0	2	Spanish/ German/ French
7	VAC	VAC101	Environmental Education-I	2	0	0	2	
8	SEC	GEN102	Principles and Applications of AI	2	0	0	2	
Total Semester Credit							**23**	

SEMESTER II								
SL. No	**Type of Course**	**Code**	**Title of the Course**	**Credit Distribution**				**Remarks**
				L	**T**	**P**	**C**	
9	CC	SDS103	Probability Distribution	3	0	1	4	CC-3
10	CC	SDS104	Sampling Theory	3	0	1	4	CC-4
11	Minor	GEN104	Basic Cyber Security and Internet of Things	4	0	0	4	
12	MDC						3	
13	AEC	AEC102	Communicative English II	2	0	0	2	
14	AEC		Foreign Language II	2	0	0	2	Spanish/ German/ French
15	VAC	VAC105	Community Engagement and Social Responsibility	1	0	1	2	
16	SEC						2	
17	SEC	CLL103	Soft Skills and Aptitude I	1	0	0	1	Soft Skills and Aptitude
Total Semester Credit							**24**	

SEMESTER III								
SL. No	**Type of Course**	**Code**	**Title of the Course**	**Credit Distribution**				**Remarks**
				L	**T**	**P**	**C**	
18	CC	SDS201	Statistical Inference	3	0	1	4	CC-5
19	CC	SDS202	Linear Algebra and Multivariate Analysis	3	0	1	4	CC-6
20	Minor						4	
21	MDC						3	
22	VAC	VAC102	Human Values and Ethics	2	0	0	2	
23	SEC						2	
24	SEC	CLL203	Soft Skills and Aptitude II	1	0	0	1	Soft Skills and Aptitude
Total Semester Credit							**20**	

SEMESTER IV								
SL. No	**Type of Course**	**Code**	**Title of the Course**	**Credit Distribution**				**Remarks**
				L	**T**	**P**	**C**	
25	CC	SDS206	Exploratory Data Analysis using Python	3	0	1	4	CC-7
26	CC	MTH208	Linear Programming and Game Theory	3	1	0	4	CC-8
27	CC	SDS207	Statistical Computing	3	0	1	4	CC-9
28	Minor						4	
29	VAC						2	
30	SEC						2	
31	SEC	CLL204	Soft Skills and Aptitude III	1	0	0	1	Soft Skills and Aptitude
Total Semester Credit							**21**	

SEMESTER V								
SL. No	**Type of Course**	**Code**	**Title of the Course**	**Credit Distribution**				**Remarks**
				L	**T**	**P**	**C**	
32	CC	SDS301	Linear Models	3	0	1	4	CC-10
33	CC	SDS30	Machine Learning	3	0	1	4	CC-11
34	CC	SDS303	Index Numbers and Time Series Analysis	3	0	1	4	CC-12
35	Minor						4	
36	SEC						2	
37	SEC	CLL303	Soft Skills and Aptitude IV	1	0	0	1	Soft Skills and Aptitude
38	INT	SDS350	Internship (Summer Project)	0	0	2	2	
Total Semester Credit							**21**	

SEMESTER VI								
SL. No	**Type of Course**	**Code**	**Title of the Course**	**Credit Distribution**				**Remarks**
				L	**T**	**P**	**C**	
39	CC	SDS304	Survey Sampling	3	0	1	4	CC-13
40	CC	SDS309	Advanced Machine Learning	3	0	1	4	CC-14
41	CC	SDS306/ SDS307	Database management System/Quantitative Methods for Finance	3	0	1	4	CC-15
42	CC	MTH311	Tools and Techniques for Scientific Report Writing	0	0	2	2	CC-16
43	Minor						4	
44	SEC						2	
45	SEC	CLL304	Soft Skills and Aptitude V	1	0	0	1	Soft Skills and Aptitude
46	Project	SDS351	Project	0	0	4	4	
Total Semester Credit							**25**	
Total Credit After 3rd Year						**134**		

SEMESTER VII								
SL. No	**Type of Course**	**Code**	**Title of the Course**	**Credit Distribution**				**Remarks**
				L	**T**	**P**	**C**	
47	CC	SDS401	Stochastic Processes and Queuing Theory	3	1	0	4	CC-16
48	CC	SDS406/ SDS403	Deep Learning/ Pricing, Hedging and Optimization	3	0	1	4	CC-17
49	CC	SDS404	Design of Experiments	3	0	1	4	CC-18
50	CC (For With research)	MTH450	Research Methodology	3	1	0	4	CC-19 (Research)
51	CC (For Without research)	SDS451/ SDS452	Statistical Quality Control / Data Analytics for Finance	3	0	1	4	CC-19 (without Research)
52	Minor						4	
53	SEC	CLL405	Soft Skills and Aptitude VI	1	0	0	1	Soft Skills and Aptitude
Total Semester Credit							**21**	

SEMESTER VIII								
S.No	**Type of Course**	**Code**	**Title of the Course**	**Credit Distribution**				**Remarks**
				L	**T**	**P**	**C**	
54	CC	SDS453	Data Engineering	3	0	1	4	CC-20
55	CC (For Without research)	SDS456/ SDS454	Image Processing and Computer Vision/Practical Financial Econometrics	3	0	1	4	CC-21 (without Research)
56	CC (For Without Research)	SDS455	Data Management and Data Warehousing	3	0	1	4	CC-22 (without Research)
57	Minor						4	
58	Minor (For without research)						4	
59	Dissertation	SDS499	Project/Dissertation	0	0	12	12	
Total Semester Credit							**20**	
Total Credit After 4th Year							**175**	

*NOTE: With research is only allowed for Students ***who secure 75% marks and above in the first six semesters***

Credit Distribution

Semester	Credit for the Following Types of Courses						
	Core Course (CC)	Minor Course	Multi-Disciplinary Course (MDC)	Ability Enhancement Course (AEC)	Value Added Course (VAC)	Skill Enhancement Course (SEC)	Project/Research Project/ Dissertation/ Internship
I	8	4	3	4	2	2	-
II	8	4	3	4	2	3	-
III	8	4	3	-	2	3	-
IV	12	4	-	-	2	3	-
V	12	4	-	-	-	3	2
VI	14	4	-	-	-	3	4
VII	16	4	-	-	-	1	-
VIII	4 (for with Research) 12 (for without Research)	4 (for with Research) & 8 (for without Research)	-	-	-	-	12 (for with Research)
Total Credit	**82 (for with Research) 90 (for without Research)**	**32 (for with Research) 36 (for without Research)**	**9**	**8**	**8**	**18**	**18 (for with Research) 6 (for without Research)**

****Total Credit attributed to Major is 94 (for with Research) and 90 (for without Research).**

Adamas University
School of Basic and Applied Sciences
Department of Mathematics

Programme Structure and Syllabus of

B.Sc (Applied Statistics and Data Science)

Programme Code: MTH3403
Duration: 4 Years Full Time
Academic Year: 2026-27

Vision of the University

To be an internationally recognized university through excellence in inter-disciplinary education, research and innovation, preparing socially responsible well-grounded individuals contributing to nation building.

Mission of the University

- Improve employability through futuristic curriculum and progressive pedagogy with cutting-edge technology
- Foster outcomes-based education system for continuous improvement in education, research and all allied activities
- Instill the notion of lifelong learning through culture of research and innovation
- Collaborate with industries, research centers and professional bodies to stay relevant and up-to-date
- Inculcate ethical principles and develop understanding of environmental and social realities

Core Values

- Respect
- Positivity
- Commitment
- Accountability
- Innovation

Vision of the School

To be recognized globally as a provider of education in Basic and Applied Sciences, fundamental and interdisciplinary research.

Mission of the School

- Develop solutions for the challenges in sciences through value-based science education.
- Conduct research leading to innovation in sciences.
- Nurture students into scientifically competent professionals in the usage of modern tools.
- Foster in students, a spirit of inquiry and collaboration to make them ready for careers in teaching, research and corporate world.

Vision of the Department

To create a Centre of academic excellence in Mathematics and Statistics through active teaching-learning and collaborative research

Mission of the Department

- Deliver graduates with considerable Mathematical and Statistical skills along with real-world problem-solving ability.
- Create a framework to nurture students through outcome-based education towards building a strong foundation in mathematical sciences for academia and industry.
- Conduct fundamental and cutting-edge collaborative research on mathematical and interdisciplinary fields.
- Contribute towards development of mathematical foundation in pan-university level.

Programme Educational Objectives (PEO) of B.Sc (Applied Statistics and Data Science):

PEO 01 Graduate will equip with latest techniques in Data Analytics like Python, Machine learning, Big Data etc.

PEO 02 Graduates will able to choose their course as a training ground to develop their positive attitude and skills.

PEO 03 Graduates of the program will become technically competent to pursue higher studies.

PEO 04 Graduates are prepared to survive in rapidly changing technology and engage in life-long learning.

PEO 05 Graduates will communicate effectively in both verbal and written form in industry and society.

Programme Outcomes (POs) and Programme Specific Outcomes (PSOs) of B.Sc (Applied Statistics and Data Science):

Students of all undergraduate general degree Programmes at the time of graduation will be able to have

PO1	**Academic Excellence**	Understanding the academic field of Statistics and its different learning areas with applications.
PO2	**Contextualized Understanding**	Develop the ability to distinguish between random and non-random experiments and simultaneously learn the theory and applications of probability.
PO3	**Design/development of solutions**	Identify, design and solve scientific problems based on data collection, data interpretation and analysis of results.
PO4	**Conduct investigations of complex problems**	Explore various real- life problems and ways to solve them with a reliable solution using various statistical methods and tests.
PO5	**Quantitative Aspects**	Learn to apply the tools of the various statistical and mathematical procedures with programming to solve real-life problems involving large data sets.
PO6	**Modernization and Tools Usage**	Develop the ability in using modern statistical, mathematical and data analytics tools for design and analysis, and quality control.
PO7	**Societal Implication**	Apply statistical methods and tools in societal, demographic, health, business and cultural issues.
PO8	**Environment and Sustainability**	Understand the tools towards problem solving and applications in biological science, agricultural science, and social sciences.
PO9	**Ethics**	Apply ethical principles and commit to professional ethics and responsibilities and norms of mathematical and data science.
PO10	**Individual and Team Work**	Work effectively as an individual or as a member or leader in undertaking projects, research organizations, industries and multidisciplinary area.
PO11	**Communication**	Build up communication skills, both written and oral, so as to apply them to write effective reports.
PO12	**Life Long Learning**	Develop the ability to evaluate theories, methods, principles, and applications of pure and applied Statistics and data science.
PSO1		Have the versatility to work effectively in a broad range of analytic, scientific, government, financial, health, technical and other positions.

PSO2		Be familiar with a variety of examples where the knowledge of mathematics or statistics helps to explain the abstract or physical phenomena accurately.
PSO3		Enhance theoretical rigor with technical skills which prepare students to become globally competitive to enter into a promising professional life in both government and private sector.

School of Basic and Applied Sciences
Department of Mathematics

B.Sc (Applied Statistics and Data Science)

Undergraduate Course Structure & Syllabuses

Course Structure

SCHOOL OF BASIC AND APPLIED SCIENCES
UNDERGRADUATE COURSE STRUCTURE
B.Sc (Applied Statistics and Data Science)

SEMESTER I								
SL. No	**Type of Course**	**Code**	**Title of the Course**	**Contact Hours Per Week**				**Remarks**
				L	**T**	**P**	**C**	
1	CC	SDS101	Probability Theory	3	1	0	4	CC-1
2	CC	SDS102	Descriptive Statistics	3	0	2	4	CC-2
3	Minor						4	
4	MDC						3	
5	AEC		English Communication I				2	
6	AEC		Foreign Language I				2	Spanish/ German/ French
7	VAC	VAC101	Environmental Education-I	2	0	0	2	
8	SEC	GEN102	Principles and Applications of AI	2	0	0	2	
Total Semester Credit							**23**	
SEMESTER II								
SL. No	**Type of Course**	**Code**	**Title of the Course**	**Contact Hours Per Week**				**Remarks**
				L	**T**	**P**	**C**	
9	CC	SDS103	Probability Distribution	3	0	2	4	CC-3
10	CC	SDS104	Sampling Theory	3	0	2	4	CC-4
11	Minor						4	
12	MDC						3	
13	AEC		English Communication II				2	
14	AEC		Foreign Language II				2	Spanish/ German/ French
15	VAC						2	
16	SEC						2	
17	SEC						1	Soft Skills and Aptitude
Total Semester Credit							**24**	
SEMESTER III								
SL. No	**Type of Course**	**Code**	**Title of the Course**	**Contact Hours Per Week**				**Remarks**
				L	**T**	**P**	**C**	
18	CC	SDS201	Statistical Inference	3	0	2	4	CC-5
19	CC	SDS202	Linear Algebra and Multivariate Analysis	3	0	2	4	CC-6
20	Minor						4	
21	MDC						3	
22	VAC	VAC102	Human Values and Ethics	2	0	0	2	
23	SEC						2	
24	SEC						1	Soft Skills and Aptitude
Total Semester Credit							**20**	

SEMESTER IV								
SL. No	Type of Course	Code	Title of the Course	Contact Hours Per Week				Remarks
				L	T	P	C	
25	CC	SDS206	Exploratory Data Analysis using Python	3	0	2	4	CC-7
26	CC	MTH208	Linear Programming and Game Theory	3	1	0	4	CC-8
27	CC	SDS207	Statistical Computing	3	0	2	4	CC-9
28	Minor						4	
29	VAC						2	
30	SEC						2	
31	SEC						1	Soft Skills and Aptitude
Total Semester Credit							**21**	
SEMESTER V								
SL. No	**Type of Course**	**Code**	**Title of the Course**	**Contact Hours Per Week**				**Remarks**
				L	**T**	**P**	**C**	
32	CC	SDS301	Linear Models	3	0	2	4	CC-10
33	CC	SDS30	Machine Learning	3	0	2	4	CC-11
34	CC	SDS303	Index Numbers and Time Series Analysis	3	0	2	4	CC-12
35	Minor						4	
36	SEC						2	
37	SEC						1	Soft Skills and Aptitude
38	INT	SDS350	Internship (Summer Project)				2	
Total Semester Credit							**21**	
SEMESTER VI								
SL. No	**Type of Course**	**Code**	**Title of the Course**	**Contact Hours Per Week**				**Remarks**
				L	**T**	**P**	**C**	
39	CC	SDS304	Survey Sampling	3	0	2	4	CC-13
40	CC	SDS309	Advanced Machine Learning	3	0	2	4	CC-14
41	CC	SDS306/ SDS307	Database management System/Quantitative Methods for Finance	3	0	2	4	CC-15
42	CC	MTH311	Tools and Techniques for Scientific Report Writing	0	0	4	2	CC-16
43	Minor						4	
44	SEC						2	
45	SEC						1	Soft Skills and Aptitude
46	Project	SDS351	Project				4	
Total Semester Credit							**25**	
Total Credit After 3rd Year						**134**		

SEMESTER VII								
SL. No	Type of Course	Code	Title of the Course	Contact Hours Per Week				Remarks
				L	T	P	C	
47	CC	SDS401	Stochastic Processes and Queuing Theory	3	1	0	4	CC-16
48	CC	SDS406/ SDS403	Deep Learning/ Pricing, Hedging and Optimization	3	0	2	4	CC-17
49	CC	SDS404	Design of Experiments	3	0	2	4	CC-18
50	CC (For With research)	MTH450	Research Methodology	3	1	0	4	CC-19 (Research)
51	CC (For Without research)	SDS451/ SDS452	Statistical Quality Control / Data Analytics for Finance	3	0	2	4	CC-19 (without Research)
52	Minor						4	
53	SEC						1	Soft Skills and Aptitude
Total Semester Credit							**21**	

SEMESTER VIII								
S.No	Type of Course	Code	Title of the Course	Contact Hours Per Week				Remarks
				L	T	P	C	
54	CC	SDS453	Data Engineering	3	0	2	4	CC-20
55	CC (For Without research)	SDS456/ SDS454	Image Processing and Computer Vision/ Practical Financial Econometrics	3	0	2	4	CC-21 (without Research)
56	CC (For Without Research)	SDS455	Data Management and Data Warehousing	3	0	2	4	CC-22 (without Research)
57	Minor						4	
58	Minor (For without research)						4	
59	Dissertation	SDS499	Project/Dissertation	12	0	0	12	
Total Semester Credit							**20**	
Total Credit After 4th Year							**175**	

*NOTE: With research is only allowed for Students ***who secure 75% marks and above in the first six semesters***

Credit Distribution

Semester	Credit for the Following Types of Courses						
	Core Course (CC)	Minor Course	Multi-Disciplinary Course (MDC)	Ability Enhancement Course (AEC)	Value Added Course (VAC)	Skill Enhancement Course (SEC)	Project/Research Project/ Dissertation/ Internship
I	8	4	3	4	2	2	-
II	8	4	3	4	2	3	-
III	8	4	3	-	2	3	-
IV	12	4	-	-	2	3	-
V	12	4	-	-	-	3	2
VI	14	4	-	-	-	3	4
VII	16	4	-	-	-	1	-
VIII	4 (for with Research) 12 (for without Research)	4 (for with Research) & 8 (for without Research)	-	-	-	-	12 (for with Research)
Total Credit	**82 (for with Research) 90 (for without Research)**	**32 (for with Research) 36 (for without Research)**	**9**	**8**	**8**	**18**	**18 (for with Research) 6 (for without Research)**

****Total Credit attributed to Major is 94 (for with Research) and 90 (for without Research).**

Syllabus

Course Title	**Probability Theory**
Course Code	**SDS101**
Credit	**4**
Contact Hours (L-T-P)	**3-1-0**
Course Type	**Theory**

Course Objectives:

The objectives of this course are as follows:

- To understand basic set theory, relations, and functions.
- To develop the concepts of probability, conditional probability, and Bayes' theorem, along with their applications.
- To gain the fundamental concepts of one and two-dimensional random variables, their properties, and applications.

Course Outcomes:

On completion of this course, the students will be able to

CO1: Recall and define fundamental concepts of sets, functions, mappings, improper integrals, infinite series, random experiments, probability axioms, random variables, and probability distributions.

CO2: Explain the principles of mathematical induction, convergence of infinite series, Beta and Gamma functions, laws of probability, conditional probability, Bayes' theorem, and properties of one- and two-dimensional random variables.

CO3: Apply mathematical induction, evaluate improper integrals including Beta and Gamma functions, test convergence of infinite series, compute probabilities using classical and axiomatic approaches, convergence, determine expectations, moments, and quantiles of random variables.

CO4: Distinguish between discrete and continuous random variables, and analyze the relationships between joint, marginal, and conditional distributions.

CO5: Assess the validity of probabilistic models, evaluate convergence criteria of series, verify properties of probability distribution functions, Bayes' theorem and total probability theorem in real-life situations.

CO6: Formulate and construct mathematical and probabilistic models involving single and two-dimensional random variables.

Course Description:

This course focus on fundamental concepts of sets, relations, and mappings, which serve as the prerequisites. The course introduces the basic principles of probability theory, focusing on one- and two-dimensional random variables. We will explore conditional probability and the application of Bayes' theorem. Additionally, the course covers both discrete and continuous random variables, examining their distribution and density functions in single and bivariate contexts. We will also delve into the expectation, dispersion, and moments of a random variable, and study the properties of two-dimensional random variables.

Course Content:

Module I: Functions, Improper integrals, Infinite series

Review of Sets, Function, Inverse function, Mapping and types of mapping, Principal of mathematical induction and its applications, Improper integrals, Beta and Gamma function, Infinite series and its convergent

Module II: Probability

Introduction, Random experiments, Sample space, Events, and algebra of events. Classical probability, Axiomatic definition of probability, Laws of addition and multiplication, Conditional probability, Independent event, Theorem of total probability, Bayes' theorem and its applications

Module III: Random variables and Probability distribution

Definition, Discrete and Continuous random variables, Probability distributions, Cumulative distribution function (c.d.f.) and its properties (with proof), Probability mass function (p.m.f.) and Probability density function (p.d.f.), expectation and moments, quantiles.

Module IV: Two-dimensional Random Variables

Discrete and continuous type, Joint, Marginal, and Conditional distributions, properties of c.d.f., Independence of random variables, theorems on sum and product of expectations of random variables

Text Books:

T1. Goon A.M., Gupta M.K. & Dasgupta B. 2002. Fundamentals of Statistics, Vol. I & II, 8th Edn. The World Press, Kolkata.

T2. Ross S. 2002. A First Course in Probability, Prentice Hall.

T3. Rohatgi V. K. and Saleh A.K. Md. E. 2009. An Introduction to Probability and Statistics. 2nd Edn. (Reprint), John Wiley and Sons.

T4. S. K. De and S. Sen A. Banerjee. 1999. Mathematical Probability, U. N. Dhur & Sons Pvt. Ltd.; 4th edition

Reference Books:

R1. Goon A.M., Gupta M.K. & Dasgupta B. 2003. An Outline of Statistical Theory, Vol I, The World Press, Kolkata.

R2. Hogg R.V., Tanis E.A. and Rao J.M. 2009. Probability and Statistical Inference, 7th Ed, Pearson Education, New Delhi.

R3. Feller W. 1968. An Introduction to Probability Theory & its Applications, John Wiley.

R4. Uspensky J.V. 1937. Introduction to Mathematical Probability, McGraw Hill.

Evaluation:

Mode of Evaluation	Theory	
Weightage	Comprehensive and Continuous Assessment	End Semester Examination
	50%	50%

Mapping COs with POs & PSOs:

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	3	1	1	1	–	–	–	–	–	1	2	2	3	2
CO2	3	3	2	2	1	–	–	–	–	–	1	2	2	3	2
CO3	3	3	3	3	2	2	1	1	–	1	1	2	3	3	3
CO4	3	3	2	2	2	1	1	1	–	1	1	2	3	3	2
CO5	3	3	3	3	2	2	2	1	1	1	1	3	3	3	3
CO6	3	3	3	3	3	3	2	1	1	2	2	3	3	3	3
	3.0	**3.0**	**2.3**	**2.3**	**1.8**	**2.0**	**1.5**	**1.0**	**1.0**	**1.3**	**1.2**	**2.3**	**2.7**	**3.0**	**2.5**

Course Title	**Descriptive Statistics**
Course Code	**SDS102**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives:
The objectives of this course are as follows:

- Introduce the basic statistics concepts, including statistical populations, samples, and data classification into different measurement scales.
- Develop skills in computing and interpreting measures of central tendency and dispersion, such as mean, median, variance, and understanding skewness and kurtosis.
- Equip students with tools to analyze relationships between variables using correlation and regression techniques, along with curve-fitting methods.
- Gain knowledge of analysing categorical data, understanding measures of association, and applying them to real-life problems involving contingency tables and attribute consistency.

Course Outcomes:
On completion of this course, the students will be able to:

CO1: Define and describe the fundamental concepts of statistics, including statistical population, types of data, levels of measurement, and graphical and tabular representations of data.

CO2: Explain and interpret measures of central tendency and dispersion, moments, skewness, and kurtosis, including real-life applications.

CO3: Apply correlation and regression techniques to bivariate data, and fit polynomial and exponential curves using the principle of least squares.

CO4: Analyze categorical data to determine independence and associations among attributes using contingency tables and measures of association like Odds Ratio, Pearson's, and Yule's measures.

CO5: Evaluate outliers and data distribution characteristics using advanced tools such as Box Plots, quantile measures, and rank correlation (e.g., Spearman's and Kendall's).

CO6: Design and execute solutions to real-world problems involving univariate, bivariate, and categorical data using statistical tools, including MS Excel and scientific calculators.

Course Description:
This course provides an introduction to statistics, covering key concepts such as types of data, scales of measurement, and methods for summarizing data through graphical and tabular representations. It explores univariate and bivariate data analysis, including measures of central tendency, dispersion, correlation, and regression. Additionally, it addresses the analysis of categorical data, contingency tables, and measures of association, with a focus on real-life applications.

Course Content:

Module I: Introduction

Definition and scope of Statistics, Population and Sample. Different types of data, Nominal, Ordinal, Rank, Interval and Ratio scales of measurement, tabular and graphical frequency distributions of data, graphical representations of data.

Module II: Univariate Data

Measures of Central Tendency: Mean, Median, Mode and related problems. Measures of Dispersion: Range, Mean deviation, Standard deviation, Quartile deviation, Coefficient of variation. Moments, Skewness and Kurtosis. Sheppard's corrections for Moments. Box Plot and Outliers detection and real-life applications.

Module III: Bivariate data

Scatter diagram, Pearson Correlation, principle of least squares, simple linear Regression, fitting of Polynomial and Exponential curves, correlation Ratio, correlation Index, Rank correlation, Spearman's and Kendall's measures, and applications in real-life problems.

Module IV: Analysis of categorical data

Theory of attributes, data consistency, Contingency table, independence and association of attributes, Measures of association - Odds ratio, Pearson's, and Yule's measure, Goodman-Kruskal gamma, and solution od real-life problems.

List of experiments (to be executed using Scientific Calculators and/or MS Excel)

1. Diagrammatic representation of data.
2. Problems based on construction of frequency distributions, cumulative frequency distributions and their graphical representations, Stem and Leaf plot.
3. Problems based on measures of Central Tendency.
4. Problems based on measures of Dispersion.
5. Problems based on combined mean and variance and coefficient of variation.
6. Problems based on Moments, Skewness and Kurtosis.
7. Problems related to Quantiles and measures based on them, construction of Box and Whisker plot.
8. Problems based on analysis of Bivariate data.
9. Problems based on measures of Rank Correlation.
10. Problems based on analysis of Categorical data

Text Books:

T1. Goon A.M., Gupta M.K. & Dasgupta B. 2002. Fundamentals of Statistics, Vol. I, 8^{th} Edn. The World Press, Kolkata.

T2. Gupta, S. C. & Kapoor, V. K. 1975. Fundamentals of Mathematical Statistics: A Modern Approach. S. Chand & Company

Evaluation:

Mode of Evaluation	Hybrid	
Weightage	Comprehensive and Continuous Assessment	End Semester Examination
	50%	50%

Mapping COs with POs & PSOs:

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	2	1	1	1	1	1	1	1	1	1	2	2	1	1
CO2	3	3	2	2	2	2	2	1	1	2	2	3	3	2	2
CO3	2	3	3	3	3	3	2	1	1	2	2	3	3	2	2
CO4	2	3	3	3	3	2	3	2	1	2	2	3	3	2	3
CO5	2	3	3	3	3	3	3	2	2	3	2	3	3	3	3
CO6	3	3	3	3	3	3	3	3	2	3	3	3	3	3	3
	2.5	**2.83**	**2.5**	**2.5**	**2.5**	**2.33**	**2.33**	**1.67**	**1.33**	**2.17**	**2**	**2.83**	**2.83**	**2.17**	**2.33**

Course Title	**Probability Distribution**
Course Code	**SDS103**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives:

- To understand the concept of different generating functions and limit Laws
- To build the knowledge of different discrete probability distributions and their applications
- To build the knowledge of different continuous probability distributions and their applications
- To explore the transformations, algebraic properties of random variables, and methods such as least squares, and regression.

Course Outcomes:

On completion of this course, the students will be able to

CO1. Recall and define the fundamental concepts of generating functions, probability inequalities, and key limit laws such as the Weak Law of Large Numbers and Central Limit Theorem.

CO2. Explain the properties and applications of discrete probability distributions, including Binomial, Poisson, and Negative Binomial distributions, and their limiting cases.

CO3. Apply the principles of continuous probability distributions, including Normal, Lognormal, and Exponential distributions, to solve real-life problems and statistical experiments.

CO4. Analyze relationships between random variables using transformations, sum and product of random variables, correlation, and regression techniques.

CO5. Evaluate real-world scenarios involving probability distributions, including fitting discrete and continuous distributions to data using computational tools like MS Excel

CO6. Design statistical models and solutions for complex real-life problems using advanced statistical methods, including Bivariate Normal Distribution, regression analysis, and least squares techniques.

Course Description:

This course provides a comprehensive introduction to probability distributions, focusing on both theoretical concepts and practical applications. It covers various probability distributions, generating functions, and probability inequalities. The course delves into key theoretical aspects of both discrete and continuous probability distributions and their practical applications in real-life scenarios. Students will gain the analytical skills needed to understand stochastic systems and apply probabilistic methods across various domains.

Course Content:

Module I: Generating Functions and Limit Laws

Probability Generating Function, Moment Generating Function and its properties, Cumulant Generating Function and Characteristic Function, Uniqueness, and Inversion theorems along with applications Markov's and Chebyshev's Inequalities, Convergence in Probability, Convergence in Mean Square and Convergence in Distribution, Weak Law of Large Numbers and their applications, De-Moivre Laplace Limit theorem, Statement of Central Limit Theorem for i.i.d. variates and its applications

Module II: Discrete Probability Distributions
Discrete Uniform, Bernoulli, Binomial, Hypergeometric, Poisson, Geometric, Negative Binomial along with their important properties and applications.

Module III: Continuous Probability Distributions
Rectangular, Normal, Lognormal, Exponential, Gamma, Beta, Cauchy, Logistic, Limiting cases, Bivariate Normal Distribution (BVN): p.d.f., properties, marginal and conditional p.d.f. of BVN

Module IV:
Transformation of Random variables, Sum and product of Random Variables, Correlation(Population) and regressions

List of experiments (to be executed using Scientific Calculators and/or MS Excel)

Sl. No.	Name of the Experiments
1	Fitting of Binomial distribution for given ‘n’ and ‘p’.
2	Fitting of Binomial distribution after computing mean and variance.
3	Fitting of Poisson distribution for given value of parameter or mean
4	Fitting of Poisson distributions after computing mean.
5	Fitting of Negative Binomial distribution.
6	Fitting of Normal distribution when parameters are given.
7	Fitting of Normal distribution when parameters are not given.
8	Fitting and application-based problems of some other Continuous distributions.
9	Application based Problems on Trinomial distribution.
10	Application based Problems on Bivariate Normal distribution.
11	Find a correlation coefficient
12	Fitting a regression line to data

Text Books

T1. Goon A.M., Gupta M.K. & Dasgupta B. 2002. Fundamentals of Statistics, Vol. I & II, 8th Edn. The World Press, Kolkata.
T2. Goon A.M., Gupta M.K. & Dasgupta B. 2003. An Outline of Statistical Theory, Vol I, The World Press, Kolkata.
T3. Ross S. 2002. A First Course in Probability, Prentice Hall.

Reference Books

R1. Rohatgi V. K. and Saleh A.K. Md. E. 2009. An Introduction to Probability and Statistics. 2nd Edn. (Reprint), John Wiley and Sons.
R2. Hogg R.V., Tanis E.A. and Rao J.M. 2009. Probability and Statistical Inference, 7th Ed, Pearson Education, New Delhi.
R3. Feller W. 1968. An Introduction to Probability Theory & its Applications, John Wiley.
R4. Uspensky J.V. 1937. Introduction to Mathematical Probability, McGraw Hill.

Evaluation:

Mode of Evaluation	Hybrid	
Weightage	Comprehensive and Continuous Assessment	End Semester Examination
	50%	50%

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	3	2	1	1	1	1	1	1	1	1	2	2	1	1
CO2	3	3	2	2	2	1	2	1	1	2	2	3	3	2	2
CO3	3	3	3	3	3	3	2	2	1	2	2	3	3	2	2
CO4	2	3	3	3	3	3	2	2	1	2	2	3	3	3	3
CO5	2	3	3	3	3	3	3	2	2	3	3	3	3	3	3
CO6	3	3	3	3	3	3	3	3	2	3	3	3	3	3	3
	2.67	**3**	**2.67**	**2.5**	**2.5**	**2.33**	**2.17**	**1.83**	**1.33**	**2.17**	**2.17**	**2.83**	**2.83**	**2.33**	**2.33**

Course Title	**Sampling Theory**
Course Code	**SDS104**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives:
The objectives of this course are as follows:

- Understand the concepts of population, parameters, and sampling distributions, and their applications.
- Explore exact sampling distributions like χ^2, t, and F, and their role in significance testing.
- Analyse bivariate normal populations and the distribution of sample statistics.
- Apply order statistics and significance testing using R-programming/MS Excel.

Course Outcomes:
On completion of this course, the students will be able to:
CO1: Define fundamental concepts such as population, parameter, random sample, statistic, sampling distributions, standard error, and transformation of variables techniques.
CO2: Explain the properties, distribution, and applications of exact sampling distributions, such as the χ^2, t-distribution, and F-distribution.
CO3: Apply sampling distributions in real-world problems, particularly in the context of the bivariate normal population and regression analysis.
CO4: Analyze the behavior of order statistics, such as the distribution of sample median, range, and joint distributions of order statistics.
CO5: Evaluate statistical tests and confidence intervals for various hypotheses, including proportions, means, variances, and regression coefficients.
CO6: Design and conduct comprehensive statistical investigations using modern statistical tools, such as R programming and MS Excel, for hypothesis testing and data analysis.

Course Description:
This course covers the theory and application of sampling distributions, exact sampling distributions (χ^2, t, and F), and their use in significance testing. Students will explore bivariate normal populations, order statistics, and various statistical tests. Practical application of these concepts will be implemented using R-programming and MS Excel.

Course Content:
Module I:
Introduction: Population and sample, statistic and sampling distribution of a statistic and its applications, sampling fluctuations & standard error, transformation of random variables and its applications.

Module II:

Exact Sampling Distributions: χ^2 distribution and its p.d.f., properties of χ^2 distribution. sampling distribution of sample mean and variance for a normal population, Student's t-distributions and its p.d.f., properties of t-distribution, F-distribution and its p.d.f., properties of F-distribution, distribution of 1/F (n1, n2), relationship between t, F and χ^2 distributions, applications of t, F and χ^2 distribution in testing of hypothesis problems.

Module III:

Bivariate normal population: Distributions of sample means, sample variances and sample correlation coefficient of a random sample from a bivariate normal population, distribution of the simple regression coefficient.

Module IV:

Order statistic: Introduction, smallest and largest order statistic, distribution of the r^{th} order statistic, joint distribution of order statistic, distribution of sample median and sample range, applications of order statistic.

List of experiments (to be executed using R-programing/MS Excel)

1. Testing of significance and confidence intervals for single proportion and difference of two proportions
2. Testing of significance and confidence intervals for single mean and difference of two means and paired tests.
3. Testing of significance and confidence intervals for difference of two variances.
4. Exact Sample Tests based on Chi-Square Distribution.
5. Testing if the population variance has a specific value and its confidence intervals.
6. Testing of independence of attributes.
7. Testing based on 2 X 2 contingency table without and with Yates' corrections.
8. Testing of significance and confidence intervals of an observed sample correlation coefficient.

Text Books

T1. Goon A.M., Gupta M.K. & Dasgupta B. 2003. An Outline of Statistical Theory, Vol I, 4th Edn., The World Press, Kolkata.

T2. Rohatgi V. K. and Saleh A.K. Md. E. 2009. An Introduction to Probability and Statistics, 2nd Edn. (Reprint), John Wiley and Sons.

Reference Books

R1. Mood A.M., Graybill F.A. and Boes D.C. 2007. Introduction to the Theory of Statistics, 3rd Edn. (Reprint), Tata McGraw-Hill Pub. Co. Ltd.

R2. Gupta S. C. & Kapoor V. K. 1975. Fundamentals of Mathematical Statistics: A Modern Approach. S. Chand & Company

R3. Hogg R.V. & Craig A.T. 1978. Introduction to Mathematical Statistics, Prentice Hall.

Evaluation:

Mode of Evaluation	Hybrid	
Weightage	Comprehensive and Continuous Assessment	End Semester Examination
	50%	50%

Mapping COs with POs & PSOs:

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	3	2	1	2	1	1	1	1	1	1	2	2	1	1
CO2	3	3	2	2	2	2	2	1	1	2	2	3	3	2	2
CO3	3	3	3	3	3	3	2	2	1	2	2	3	3	2	2
CO4	2	3	3	3	3	3	2	2	1	2	2	3	3	3	3
CO5	3	3	3	3	3	3	3	2	2	3	3	3	3	3	3
CO6	3	3	3	3	3	3	3	3	2	3	3	3	3	3	3
	2.83	**3**	**2.67**	**2.5**	**2.67**	**2.5**	**2.17**	**1.83**	**1.33**	**2.17**	**2.17**	**2.83**	**2.83**	**2.33**	**2.33**

Course Title	**Statistical Inference**
Course Code	**SDS201**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives:

The objectives of this course are as follows:

- Develop a strong foundation in statistical inference, including estimation theory and hypothesis testing with their theoretical properties.
- Understand and apply classical parametric tests and nonparametric methods for analyzing real-world data.
- Gain conceptual clarity in advanced inferential techniques such as Neyman–Pearson lemma, likelihood ratio tests, Rao-Blackwell and Lehmann-Scheffé theorems.

Course Outcomes:

On completion of this course, the students will be able to:

CO1: Apply methods of point estimation and interval estimation

CO2: Explain the theoretical foundations of estimation theory

CO3: Understand the fundamental concepts of hypothesis testing

CO4: Analyse theoretical aspects of hypothesis testing, including Neyman–Pearson lemma

CO5: Construct uniformly most powerful unbiased tests, LRT and power curve

CO6: Compare nonparametric tests with parametric methods for appropriate data analysis

Course Description:

This course introduces the fundamental concepts of statistical inference, including estimation theory and hypothesis testing. It covers properties of estimators, parametric tests, and theoretical results such as Neyman–Pearson lemma and likelihood ratio tests. The course also includes nonparametric methods, enabling students to apply appropriate statistical techniques to real-world data.

Course Content:

Module I: Estimation

Basics of Estimation theory: Introduction, methods of point estimation, interval estimation, and its applications, and mean square error, unbiasedness, best linear unbiasedness and minimum variance unbiasedness, uniformly minimum variance unbiased estimators, consistent estimators and asymptotic efficiency, sufficiency, factorization theorem (discrete case only), Fisher's information (for single parameter only). Cramer-Rao inequality and minimum variance bound estimators, Rao-Blackwell and Lehmann-Scheffe theorems and applications.

.

Module II: Hypothesis

Hypothesis Testing: Null and alternative hypotheses, simple and composite hypotheses, critical region, type I and type II errors, level of significance, size, power, p-value. exact tests and confidence intervals. Test for binomial proportions, test for Poisson mean, test for means and variances from Gaussian distribution, chi-square test for goodness of fit and independence, and test for correlation coefficient.

Module III: Hypothesis testing

Theory of hypothesis testing: Test function, randomized and non-randomized tests, most powerful test, uniformly most powerful test, Neyman-Pearson lemma and its applications, uniformly most powerful unbiased tests, likelihood ratio tests.

Module III: Nonparametric Inference

Introduction to nonparametric tests: Concept of nonparametric tests and their advantages over parametric tests. Run test for randomness, Kolmogorov Smirnov test, Sign test, and Wilcoxon signed-rank test, Mann Whitney U test and Kruskal Wallis test. Applications and comparison with corresponding parametric tests.

List of experiments (to be executed using Scientific Calculators and/or R Programming)

Sl. No.	**Name of the Experiments**
1.	Point Estimation and Properties of Estimators
2.	Estimation by the Method of Moments.
3.	Maximum Likelihood Estimation.
4.	Test of significance for single proportion and difference of two proportions.
5.	Test of significance for single Poisson mean and difference of two Poisson means.
6.	Test of significance and confidence intervals for single mean and difference of two means.
7.	Test of significance and confidence intervals for single variance and ratio of two variances.
8.	Test of parameters under Bivariate Normal distribution.
9.	Type I and Type II Errors, Most Powerful Critical Region (Neyman–Pearson Lemma) and Power Curves
10.	Run test, Sign test, Wilcoxon signed-rank test
11.	Mann-Whitney U test and Kruskal-Walli's test.

Text Books:

T1. Rohatgi V. K. and Saleh, A.K. Md. E. 2009. An Introduction to Probability and Statistics. 2ndEdn. (Reprint) John Wiley and Sons

T2. Casella G. and Berger R.L. 2002. Statistical Inference, 2nd Edn., Thomson Learning.

T3. Gupta S. C., & Kapoor V. K. 1975. Fundamentals of Mathematical Statistics: A Modern Approach. S. Chand & Company

Reference Books:

R1. Goon A.M., Gupta M.K. & Dasgupta B. 2005. An Outline of Statistical Theory, Vol. I & II, The World Press, Kolkata

R2. Miller I. and Miller M. 2002. John E. Freund's Mathematical Statistics (6th addition, low price edition), Prentice Hall of India.

Evaluation:

Mode of Evaluation	Hybrid	
Weightage	Comprehensive and Continuous Assessment	End Semester Examination
	50%	50%

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	2	2	2	2	1	2	1	2	2	1	2	3	2	3
CO2	3	3	2	2	2	2	2	2	2	2	2	3	3	3	3
CO3	2	3	3	3	3	2	3	2	3	3	2	3	3	3	3
CO4	2	3	3	3	3	3	3	2	3	3	2	3	3	3	3
CO5	3	2	3	3	3	3	3	3	2	2	2	2	3	3	3
CO6	3	2	3	3	3	3	3	2	2	3	3	3	3	3	3
	2.67	**2.5**	**2.67**	**2.67**	**2.67**	**2.33**	**2.67**	**2**	**2.33**	**2.5**	**2**	**2.67**	**3**	**2.83**	**3**

Course Title	**Linear Algebra and Multivariate Analysis**
Course Code	**SDS202**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives:
Equip students with the skills to solve linear equations and diagonalize matrices, understand concepts of basis and dimension, analyze multivariate data and distributions, and apply dimension reduction techniques and factor analysis.

Course Outcomes:
On completion of this course, the students will be able to

CO1: Recall and define the key concepts of matrix operations, echelon forms, matrix inverse, rank of a matrix, and basic properties of eigenvalues and eigenvectors.

CO2: Explain the theory behind linear systems, vector spaces, subspaces, linear dependence and independence, and the significance of eigenvalues and eigenvectors in the context of diagonalization and spectral decomposition.

CO3: Apply matrix operations and eigenvalue decomposition methods to solve linear equations, compute matrix inverses, and perform diagonalization and spectral decomposition of matrices.

CO4: Analyze and solve multivariate problems, such as computing mean vectors, dispersion matrices, and performing marginal and conditional distributions.

CO5: Critically evaluate the effectiveness of multivariate analysis techniques such as discriminant analysis, principal component analysis (PCA), and factor analysis in various applications, including the extraction of key features from data and classification.

CO6: Design and implement multivariate statistical methods such as discriminant analysis, principal components analysis, and factor analysis for solving complex real-world problems.

Course Description:
This is a foundational course in mathematics and statistics that covers key topics such as solving systems of linear equations, matrix diagonalization, understanding basis and dimension, linear mappings, inner products, multivariate data, multivariate distributions, sampling distribution, dimension reduction techniques, and factor analysis. This course provides students with essential theoretical knowledge and practical tools to analyze multivariate data and make informed inferences. The course includes lectures, assignments, quizzes, and practical exercises to help students grasp and apply the concepts of linear algebra and multivariate analysis in real-world scenarios. By the end of the course, students will be equipped with the skills to analyze complex multivariate data and draw meaningful conclusions for decision-making.

Course Content:

Module I

Elementary row operations, echelon forms, matrix inverse, rank of matrix, solution to the system of linear equations, applications of linear equations, Cayley-Hamilton theorem, eigen value and eigen vector, diagonalization, spectral decomposition theorem, quadratic form

Module II

Vector spaces, Subspaces, Linear dependence and independence, dimension and basis, linear mapping, rank-nullity theorem and its applications, Inner Product, orthogonality of vectors,

Module III

Multivariate data: Random vector, probability mass and density function, distribution function, mean vector & dispersion matrix, marginal & conditional distributions, multivariate normal distribution and its properties, sampling distribution for mean vector and variance-covariance matrix

Module IV

Applications of multivariate analysis: discriminant analysis, principal components analysis and factor analysis

List of Practical using R:

1. Bivariate Normal Distribution,
2. Multivariate Normal Distribution
3. Discriminant Analysis
4. Principal Components Analysis
5. Factor Analysis
6. Finding eigen value and eigen vector of a variance-covariance matrix

Text Books:

T1. Anderson, T.W. 2003. An Introduction to Multivariate Statistical Analysis, 3rdEdn., John Wiley

T2. Johnson, R.A., and Wichern, D.W. 2007. Applied Multivariate Analysis, 6thEdn., Pearson & Prentice Hall

T3. S. K. Mapa. 2003. Higher Algebra- Abstract and Linear, revised Ninth Edition, Sarat Book House.

Reference Books:

R1. Muirhead, R.J. 1982. Aspects of Multivariate Statistical Theory, John Wiley

R2. S D Sharma, Operations Research: Theory and Applications, 4th edition, Laxmi Publications

R3. Kshirsagar, A.M. 1972. Multivariate Analysis, 1stEdn. Marcel Dekker

R4. Mukhopadhyay, P. 1996. Mathematical Statistics, New Central Book Agency.

R5. Gibbons, J. D. and Chakraborty, S. 2003. Nonparametric Statistical Inference. 4th Edition. Marcel Dekker, CRC.

Course Title	Exploratory Data Analysis using Python with Python
Course Code	**SDS206**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives:

The objectives of this course are as follows:

- To make students understand the mathematical foundations of programming, Python execution model, data structures, and algorithm design principles.
- To enable students to apply scientific computing techniques, NumPy operations, and visualization tools for structured data analysis.
- To develop students' proficiency in designing efficient algorithms.
- To analyze high-dimensional datasets using exploratory data analysis, covariance structures, PCA concepts, and bias detection techniques.
- To develop the ability to create reproducible analytical pipelines integrating feature engineering, advanced visualization, and scalable data processing methods.

Course Outcomes:

On completion of this course, the students will be able to:

CO1: Recall fundamental concepts of Python programming, mathematical foundations, core data structures, algorithm design principles.

CO2: Explain the Python execution, object-oriented concepts, algorithmic complexity, matrix algebra foundations feature representation techniques.

CO3: Apply Python programming, NumPy operations, data structures, searching/sorting algorithms, scientific visualization tools, and EDA.

CO4: Analyze algorithm efficiency, high-dimensional data structures, covariance and correlation patterns, PCA-based transformations, bias detection mechanisms.

CO5: Evaluate scaling strategies, visualization integrity, fairness metrics, and feature importance using appropriate computational and statistical criteria.

CO6: Create Python-based analytical pipelines integrating data structures, scientific computing, exploratory data analysis, advanced visualization, and feature engineering for high-dimensional datasets.

Course Description:

This course builds a rigorous foundation in Python programming, scientific computing, and data-driven problem solving, preparing students for advanced machine learning, deep learning, and high-performance computing. Learners will master core programming concepts, memory-efficient data structures, algorithmic thinking, and performance-aware workflows, alongside high-dimensional data analysis, statistical geometry, and reproducible, bias-aware visualization. By integrating mathematical rigor with practical computational skills, the course equips students to tackle real-world, large-scale datasets and lays the groundwork for cutting-edge ML/AI research, including deep learning pipelines.

Course Content:

Module I: Mathematical Foundations of Programming

Python Execution Model, Data Types and Memory Management, Control Flow and Functions, Recursion and Modular Programming, Object-Oriented Concepts, Exception Handling,

Core Data Structures: Lists, Tuples, Sets, Dictionaries, Stacks, Queues, Linked Lists, Hash Tables, Trees and Graph Representations.

Algorithm Design: Searching and Sorting, Divide and Conquer, Basic Dynamic Programming, Asymptotic Complexity Analysis (Big-O).

NumPy: Vectorization, Multi-Dimensional Arrays, Matrix and Vector Operations, Floating-Point Arithmetic, Numerical Stability.

Module II: Scientific Computing, Data Visualization and DSA Concepts

High-Dimensional Array Computation: Matrix Algebra, Eigenvalue Decomposition, Sparse Data Structures.

Advanced Data Structures: Heaps, Balanced Trees; Graph Traversal: BFS, DFS; Shortest Path Concepts, Scaling Principles.

Scientific Data Visualization: Matplotlib, Seaborn, Univariate & Multivariate Plots, Correlation & Covariance Heatmaps.

High-Dimensional Visualization: Projection Plots, Image Data as Matrices, Visualization Bias.

Module III: Exploratory Data Analysis and Visualization

Foundations of EDA, Distribution Diagnostics, Robust Descriptive Statistics, Outlier, Leverage & Influence Detection, Covariance Structure Analysis, PCA.

Correlation Geometry & Dependence Mapping, Structural Data Transformations, Scaling & Normalization, Log/Rank Transform Techniques. Bias Detection, Multi-Dimensional Projection Visualization, Sequential Data Analysis, Missing Pattern Diagnostics.

Module IV: Advanced Visualization and Feature Representation Engineering

Feature Construction, Scaling Effects, Interaction Effects Visualization, Sparse Feature Exploration, High-Dimensional Covariance Mapping, Multi-Feature Scatterplots, Feature Ranking & Exploratory Importance Assessment, Structural Transformations for Visualization (PCA, Rotation), Covariance and Correlation Heatmaps.

List of Experiments using Python/R

1. Student Record Manager using Python Data Structures: Design and implement a Python program to manage student records using core data structures such as lists and dictionaries. The program should allow operations like adding, updating, deleting, and searching student details. Analyze the time complexity of search operations and demonstrate exception handling for invalid inputs.
2. Recursive Problem Solver and Function Modularization: Develop a Python program to solve problems like factorial computation and Fibonacci series using recursion.

Compare recursive and iterative approaches in terms of efficiency and memory usage. Structure the program using modular functions and analyze the call stack behavior.

3. Sorting and Searching Algorithm Visualizer: Implement basic sorting algorithms (Bubble Sort, Selection Sort) and searching techniques (Linear Search, Binary Search) in Python. Compare their performance using different input sizes and analyze their time complexity using Big-O notation. Optionally visualize the step-by-step process.
4. Matrix Operations using NumPy: Create a Python program using the NumPy library to perform matrix and vector operations such as addition, multiplication, transpose, and inversion. Demonstrate the advantages of vectorization over traditional loops and discuss numerical stability and floating-point precision issues.
5. High-Dimensional Matrix Operations and Eigenvalue Analysis: Develop a Python program using NumPy to perform matrix algebra operations such as multiplication, transpose, and eigenvalue decomposition. Demonstrate how high-dimensional arrays are handled efficiently and interpret the significance of eigenvalues in data analysis.
6. Graph Traversal and Shortest Path Implementation: Implement graph traversal techniques including Breadth-First Search (BFS) and Depth-First Search (DFS) using adjacency list representation. Extend the program to compute the shortest path in a graph (unweighted or basic weighted case) and analyze performance for scalability.
7. Data Visualization and Correlation Heatmap Analysis: Using Matplotlib and Seaborn, create univariate and multivariate visualizations such as histograms, scatter plots, and pair plots. Generate correlation and covariance heatmaps and interpret relationships between variables.
8. Exploratory Data Analysis and Distribution Diagnostics: Perform exploratory data analysis (EDA) on a real-world dataset using Pandas and Seaborn. Analyze data distributions using histograms, box plots, and density plots. Identify skewness, detect outliers, and compute robust descriptive statistics such as median and interquartile range.
9. Data Transformation and Normalization Techniques: Implement data preprocessing techniques including scaling, normalization, log transformation, and rank transformation using NumPy. Compare the effects of these transformations on skewed datasets and evaluate improvements in data distribution and model readiness.
10. Correlation Analysis and Principal Component Analysis (PCA): Compute and visualize correlation matrices and covariance structures using NumPy and Seaborn. Apply Principal Component Analysis (PCA) to reduce dimensionality and visualize high-dimensional data in 2D space. Interpret variance explained by principal components.

Evaluation:

Mode of Evaluation	**Theory**		**Practical**	
Weightage	Comprehensive and Continuous Assessment	End Semester Examination	Comprehensive and Continuous Assessment	End Semester Examination
	25%	25%	25%	25%

Text Books:

T1. Michael T. Goodrich, Roberto Tamassia, Michael H. Goldwasser (2021), Data Structures and Algorithms in Python, Wiley India edition

T2. Wes McKinney (2022), Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Jupyter, O'Reilly Media, Inc, 3rd Edition

Reference Books:

R1. Ellis Horowitz, Sartaj Sahni, Sanguthevar Rajasekaran (2008), Fundamentals of Computer Algorithms, Universities Press (India), 2nd Edition.

Evaluation:

Mode of Evaluation	Hybrid			
Weightage	Comprehensive and Continuous Assessment		End Semester Examination	
	Theory	Practical	Theory	Practical
	25%	25%	25%	25%

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	2	1	1	2	2	1	1	2	2	1	1	2	2	3
CO2	3	3	2	2	2	2	2	1	2	2	2	2	3	3	3
CO3	3	2	3	2	3	3	3	2	3	3	2	3	3	3	3
CO4	2	3	3	3	3	3	3	2	3	3	2	3	3	3	3
CO5	2	2	3	3	3	3	3	3	2	3	2	2	3	3	3
CO6	3	2	3	3	3	3	3	3	2	3	3	3	3	3	3
	2.67	**2.33**	**2.50**	**2.33**	**2.67**	**2.67**	**2.50**	**2.00**	**2.33**	**2.67**	**2.00**	**2.33**	**2.83**	**2.83**	**3.00**

Course Title	Statistical Computing
Course Code	SDS207
Credit	4
Contact Hours (L-T-P)	3-0-2
Course Type	Hybrid

Course Objectives:

The objectives of this course are as follows:

1. To develop strong foundations in simulation techniques and random variate generation for modeling stochastic systems.
2. To enable implementation of Monte Carlo integration and optimization with variance reduction and convergence acceleration.
3. To understand efficiency and convergence properties of Monte Carlo estimators and related methods.
4. To build competence in resampling techniques and missing data methods including EM and MM algorithms.
5. To develop skills in statistical validation and modeling for reliable inference under uncertainty.

Course Outcomes:

On completion of this course, the students will be able to:

CO1: Recall simulation fundamentals, including random number generation and sampling methods for discrete, continuous, mixture, spherical, and convex polygon distributions

CO2: Explain Monte Carlo integration, importance sampling, and convergence control techniques including asymptotic variance and optimization methods

CO3: Apply Monte Carlo methods to generate samples, estimate integrals, and improve efficiency using variance reduction techniques

CO4: Analyze simulated data using jackknife, bootstrap, cross-validation, and implement EM and MM algorithms

CO5: Evaluate statistical validation and variance reduction techniques for accuracy and reliability of results

CO6: Develop integrated simulation-based solutions for stochastic problems using Monte Carlo and optimization methods

Course Description:

This course introduces students to fundamental and advanced techniques in simulation and random variate generation, including inversion, rejection, and composition methods for discrete, continuous, mixture, spherical, and convex polygon distributions. It covers classical Monte Carlo integration methods, importance sampling, and techniques for controlling and accelerating convergence, including variance analysis and optimization of Monte Carlo estimators. The course also emphasizes statistical analysis of simulated data through resampling methods such as jackknife and bootstrap, along with the formulation and application of EM and MM algorithms. Additionally, it includes variance reduction techniques, statistical validation methods, and cross-validation approaches to ensure accurate and reliable inference in stochastic modeling.

Course Content:

Module I:
Random number generation, Inversion method, Rejection Principle, Composition Approach, Random sample generation from discrete distributions, continuous distributions, convex polygon, mixture distributions, spherical distribution

Module II:
Classical Monte Carlo Integration techniques, Importance Sampling, Controlling and Accelerating Convergence: Asymptotic variance of importance sampling estimators, Acceleration Methods, Numerical Optimization Methods, Optimizing Monte Carlo approximation.

Module III:
Statistical analysis of simulated data, Jackknife and Bootstrap, Formulation of the EM and MM algorithms with illustrations, Variance reduction techniques, Statistical validation techniques, Cross validation

List of Experiments using R

1. **Random Number Generation:** Implement pseudo-random number generators (such as Linear Congruential Generator) to produce uniform random numbers and analyze their statistical properties like uniformity, independence, mean, and variance
2. **Generation of Discrete Random Variables:** Use the inversion method to generate random samples from discrete probability distributions and verify that the empirical frequencies match the theoretical probabilities
3. **Generation of Continuous Random Variables** Apply inversion and rejection methods to generate samples from continuous distributions and validate their correctness using graphical (histogram) and statistical comparisons with theoretical distributions
4. **Generation of Random Variables from Mixture, Spherical, and Convex Polygon Distributions:** Use composition method to generate samples from mixture distributions, uniformly distribute points on a sphere, and generate random points inside a convex polygon, then verify their distributional properties
5. **Monte Carlo Integration:** Estimate definite integrals using classical Monte Carlo techniques by generating uniform random samples over a domain and approximating the integral as an average of function values, then compare the result with the exact value to study accuracy and convergence
6. **Importance Sampling for Variance Reduction:** Implement importance sampling to estimate integrals more efficiently by choosing an appropriate proposal distribution, and compare the variance and accuracy of the estimator with standard Monte Carlo methods.
7. **Optimization and Convergence Acceleration in Monte Carlo Methods:** Apply basic numerical optimization techniques to improve Monte Carlo estimators, and use

acceleration methods to reduce error and enhance convergence speed, analyzing the effect on variance

8. **Bootstrap and Jackknife Methods for Statistical Analysis:** Apply bootstrap and jackknife resampling techniques to simulated data to estimate parameters (mean, variance) and their standard errors, and compare the results with theoretical values to assess accuracy
9. **Implementation of EM Algorithm for Parameter Estimation:** Formulate and implement the Expectation-Maximization (EM) algorithm on simulated incomplete data (e.g., mixture distribution) to estimate unknown parameters, and analyze convergence behavior through iterations
10. **Cross-Validation and Variance Reduction Techniques:** Use cross-validation methods to evaluate model performance on simulated data and apply variance reduction techniques (such as control variates or antithetic variates) to improve estimation efficiency, comparing results before and after improvement.

Evaluation:

<table>
<tr><th>Mode of Evaluation</th><th colspan="2">Theory</th><th colspan="2">Practical</th></tr>
<tr><td rowspan="2">Weightage</td><td>Comprehensive and Continuous Assessment</td><td>End Semester Examination</td><td>Comprehensive and Continuous Assessment</td><td>End Semester Examination</td></tr>
<tr><td>25%</td><td>25%</td><td>25%</td><td>25%</td></tr>
</table>

Text Books:

T1. Christian P. Robert, George Casella, Introducing Monte Carlo Methods with R, Springer

T2. Sheldon Ross, Simulation, Elsevier AP, 5th Edition.

Reference Books:

R1. Peter J. Huber & Elvezio M. Ronchetti, Robust Statistics, Wiley (Wiley Series in Probability and Statistics), 2nd Edition (2009)

R2. Christian P. Robert & George Casella, Monte Carlo Statistical Methods, Springer New York (Springer Texts in Statistics), 2nd Edition (2004).

Course Title	**Linear Programming and Game Theory**
Course Code	**MTH208**
Credit	**4**
Contact Hours (L-T-P)	**3-1-0**
Course Type	**Theory**

Course Objectives:

The objectives of this course are as follows:

- To develop the concept of formulation of linear programming problem (LPP) and solution procedure of LPP by using Graphical method, simplex method.
- To acquire the knowledge of special classes of LPPs such as Transportation problem, Assignment problem and Travelling salesman problem
- To build up the concept of application procedures of LPP to Game Theory.

Course Outcomes:

On completion of this course, the students will be able to

CO1: Define key concepts of LPP, simplex, transportation, assignment, and game theory

CO2: Explain the methods and basic principles of LPP, transportation, assignment, and game theory

CO3: Apply simplex method, transportation and assignment problem solving methods, and game theory

CO4: Analyze solutions in LPP, duality theory, and optimization methods for transportation and assignment problems

CO5: Evaluate methods for solving LPP, transportation, assignment problems, and game theory problems

CO6: Design and create complex models for LPP, transportation, assignment problems, and games

Course Description:

Linear Programming Problem (LPP) deals with the problem of minimizing or maximizing a linear function in the presence of linear equalities/inequalities. Since the development of the simplex method, linear programming has been extensively used in the military, industrial, governmental, and urban planning fields, among others. A large class of optimization problems are solved by techniques under the heading mathematical programing. There are several other techniques to solve the optimization problems but LPP concerned with only solving the linear programing models in an iterative procedure which yields an exact solution in a finite number of steps. LPP is useful to solve the allocation problems. These problems require the allocation of limited available resources to the jobs that are to be done. There is a class of games which are intimately related to linear programing.

Course Content:

Module I:

Review of linear programming problem (LPP) formation, convex set & Convex function, convex hull and convex polyhedron, Extreme points, Hyperplane, Supporting and separating hyperplane, Basic solutions and basic feasible solution (BFS), Degenerate and non-degenerate BFS, Reduction of a feasible solution (FS) to a BFS, Optimality condition, Unboundedness, Alternate optima, Infeasibility

Module II:
Theory of simplex method, the simplex algorithm, optimality and un-boundedness, two-phase method, Big-M method, Duality, formulation of the dual problem, primal-dual relationships, Duality theorem, interpretation of the dual economically

Module III:
Transportation problem and its mathematical formulation, North-West corner method, Least cost method and Vogel approximation method for determination of initial basic feasible solution, Optimality test, Modified Distribution (MODI) method; Assignment problem and its mathematical formulation, Hungarian method for solving an assignment problem; Travelling salesman problem (TSP)

Module IV:
Concept of game theory, rectangular games, pure strategy and mixed strategy, saddle point, optimal strategy and value of the game, concept of dominance, formulation of two-person zero-sum games, solving two-person zero-sum games, games with mixed strategies, graphical solution procedure, linear programming solution of games

Evaluation:

Mode of Evaluation	Theory	
Weightage	Comprehensive and Continuous Assessment	End Semester Examination
	50%	50%

Text Books:

T1. Kanti Swarup, P. K. Gupta and Man Mohan, Operations Research, S. Chand and Co. Pvt. Ltd.
T2. Hamdy A. Taha, Operations Research, An Introduction, 8th Ed., Prentice Hall India, 2006.
T3. G. Hadley, Linear Programming, Narosa Publishing House, New Delhi, 2002.

Reference Books:

R1. F.S. Hillier and G.J. Lieberman, Introduction to Operations Research, 9th Ed., Tata McGraw Hill, Singapore, 2009.
R2. N.V.R. Naidu, G. Rajendra and T. Krishna Rao-Operations Research, I.K. International Publishing House Pvt. Ltd., New Delhi, Bangalore.
R3. D. C. Sanyal, K. Das, Linear Programming and Game Theory, U. N. Dhur and sons Pvt Ltd, 2008.

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	3	2	1	1	1	1	1	2	1	1	1	3	2	2
CO2	3	3	2	2	1	1	1	1	2	1	1	1	3	2	2
CO3	3	3	3	2	2	2	1	1	3	2	2	2	3	3	3
CO4	3	3	3	3	2	2	2	2	3	2	3	2	3	3	3
CO5	3	3	3	3	3	3	2	2	3	3	3	3	3	3	3
CO6	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	3	**3**	**2.67**	**2.33**	**2**	**2**	**1.67**	**1.67**	**2.67**	**2**	**2.17**	**2**	**3**	**2.67**	**2.67**

Course Title	**Data Analytics Principles and Techniques**
Course Code	**SDS205**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives:
Equip students with foundational knowledge and practical skills in data analytics, big data, machine learning, and real-time analytics, while emphasizing ethical considerations in data science practices.

Course Outcomes:
On completion of this course, the students will be able to:
CO1: Define the key concepts and processes in data analytics, big data, and machine learning, emphasizing phases, challenges, and tools.
CO2: Explain the role of data preprocessing, visualization, and exploratory data analysis in building insights and improving data quality.
CO3: Apply machine learning models, data preprocessing, and visualization techniques to solve real-world problems using Python or R.
CO4: Analyze the performance of clustering, association rule mining, and scalable algorithms to derive meaningful insights from datasets.
CO5: Evaluate the ethical implications of data science practices, including privacy, diversity, and inclusion, in case studies and real-time applications.
CO6: Design and develop advanced analytical and machine learning workflows, including scalable clustering and real-time analytics, for complex datasets.

Course Description:
This course provides a comprehensive exploration of data analytics, big data, and machine learning techniques, focusing on data visualization, real-time analytics applications, and ethical considerations in data science, equipping students with the tools to analyze and interpret complex data effectively.

Course Content:
Module I
Data Analytics - Types – Phases - Quality and Quantity of data – Measurement - Exploratory data analysis - Business Intelligence. Introduction to data visualization – Data visualization options – Filters – Dashboard development tools

Module II
Introduction to big data: Introduction to Big Data Platform – Challenges of Conventional Systems - Intelligent data analysis – Nature of Data - Analytic Processes and Tools - Analysis vs Reporting. Mining data streams: Introduction to Streams Concepts – Stream Data Model and Architecture - Stream Computing - Sampling Data in a Stream – Filtering Streams – Counting Distinct Elements in a Stream.

Module III

Machine learning – Modeling Process – Training model – Validating model – Predicting new observations –Supervised learning algorithms – Unsupervised learning algorithms. Graphical and sequential models- Bayesian networks- conditional independence- Markov random fields- inference in graphical models- Belief propagation- Markov models- Hidden Markov models- decoding states from observations- learning HMM parameters.

Module IV

Real time Analytics Platform (RTAP) Applications - Case Studies - Real Time Sentiment Analysis- Stock Market Predictions. Data Science Ethics – Doing good data science – Owners of the data - Valuing different aspects of privacy - Getting informed consent - The Five Cs – Diversity – Inclusion – Future Trends.

Practical/Lab to be performed on a computer using R/PYTHON programming:

1. Implement Data preprocessing methods on student and labour datasets and data cube for data warehouse on 3-dimensional data
2. Implement various missing handling mechanisms, Implement various noisy handling mechanisms
3. Develop k-means and MST based clustering techniques, Develop the methodology for assessment of clusters for given dataset
4. Design algorithms for association rule mining algorithms
5. Derive the hypothesis for association rules to discovery of strong association rules; Use confidence and support thresholds.
6. Construct Heat wavelet transformation for numerical data, Construct principal component analysis (PCA) for 5-dimensional data.
7. Implement binning visualizations for any real time dataset, Implement linear regression techniques
8. Visualize the clusters for any synthetic dataset, Implement the program for converting the clusters into histograms
9. Write a program to implement agglomerative clustering technique. Write a program to implement divisive hierarchical clustering technique
10. Develop scalable clustering algorithms, Develop scalable a priori algorithm.

Text Books

T1. Jiawei Han, Micheline Kamber and Jian Pei. 2011. Data Mining Concepts and Techniques, Third Edition, Elsevier.

T2. Michael Berthold, David J. Hand, 2007. Intelligent Data Analysis, Springer.

T3. Tom White, 2012. Hadoop: The Definitive Guide, Third Edition, O’reilly, Media.

T4. Chris Eaton, Dirk DeRoos, Tom Deutsch, George Lapis, Paul Zikopoulos, 2012. Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, McGrawHill Publishing.

T5. Anand Rajaraman and Jeffrey David Ullman, “Mining of Massive Datasets”, CUP,2012.

Reference Books

R1. Ramez Elmasri and Shamkant B. Navathe. Fundamentals of Database Systems, 7th edition, Addison Wesley, 2015.

R2. Anil Maheshwari. Data Analytics Made Accessible, Amazon Digital Services LLC, 2014.
R3. Jure Leskovec, Anand Rajaraman and Jeffrey David Ullman. Mining of Massive Datasets, 2nd edition, Cambridge University Press, 2014.
R4. Jiawei Han, Micheline Kamber, Jian Pei. Data Mining: Concepts and Techniques, 3rd edition, Morgan Kaufmann, 2011.
R5. Knaflic, Cole N. Storytelling with Data: a Data Visualization Guide for Business Professionals. Hoboken, New Jersey: Wiley, 2015.
R6. Stephanie Evergreen. Effective Data Visualization: The Right Chart for the Right Data, SAGE Publications, Inc, 1st edition, 2016.
R7. Nataraj Venkataramanan and Ashwin Shriram. Data Privacy: Principles and Practice, 1st edition, Chapman and Hall/CRC, 2016.

Evaluation:

Mode of Evaluation	Hybrid			
Weightage	Comprehensive and Continuous Assessment		End Semester Examination	
	Theory	Practical	Theory	Practical
	25%	25%	25%	25%

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	2	1	1	2	1	1	1	1	1	1	2	2	2	1
CO2	3	3	2	2	3	2	1	1	1	1	2	2	3	3	2
CO3	2	2	3	2	3	3	1	1	2	2	2	3	3	3	3
CO4	3	2	3	3	3	3	1	2	1	2	2	3	3	3	3
CO5	2	1	2	2	2	2	3	3	3	1	2	2	2	2	2
CO6	3	2	3	3	3	3	2	2	2	3	3	3	3	3	3
	2.67	**2.00**	**2.33**	**2.17**	**2.67**	**2.33**	**1.50**	**1.67**	**1.67**	**1.67**	**2.00**	**2.50**	**2.67**	**2.67**	**2.33**

Course Title	**Linear Models**
Course Code	**SDS301**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives:
To introduce the students with the method of least squares and Linear Models both Simple and Multiple Regression Models, ANOVA Models and finally use some tools to test the model validity

Course Outcomes:
On completion of this course, the students will be able to
CO1: Understand and apply the Gauss-Markov theorem for linear estimation, including error variance estimation and least squares methods.
CO2: Perform simple and multiple regression analysis, including hypothesis testing and model validation.
CO3: Use Bonferroni and Scheffe's methods for simultaneous interval estimation under linear restrictions.
CO4: Apply least squares methods and understand orthogonal splitting of variation in linear models.
CO5: Conduct ANOVA for fixed and random effect models with one-way and two-way classified data.
CO6: Identify violations in regression assumptions and apply logistic and Poisson regression for binary and count data using GLM concepts.

Course Description:
This course provides a detailed examination of advanced statistical methods, covering linear estimation, regression analysis, ANOVA, and the handling of violations in regression assumptions, including binary and count data techniques.

Course Content:
Module I: Gauss-Markov Set-up
Theory of Linear Estimation, Estimability of linear parametric functions, Gauss-Markov theorem, Estimation of error variance. Fundamental Theorems on Linear Models, BLUE, Orthogonal Splitting of total variation, selection of Valid Error. Estimations under linear restriction on parameters. Simultaneous Interval estimation (Bonferroni t, Scheffe's S Method)

Module II: Regression Analysis
Simple and Multiple Regression analysis. Significance of regression parameters. Tests for Parallelism and Identity, Linearity of Simple Regression. Model checking: Prediction from a fitted model. Concept of Moore Penrose G- Inverse.

Module III: Analysis of Variance

Definitions of Fixed, Random and Mixed Effect Models, Analysis of Variance in One-way classified data for Fixed Effect Models, Analysis of Variance in Two-way classified data with equal number of observations per cell for Fixed Effect Models. Analysis of Variance in One-way classified data for Random Effect Models.

Module IV: Violations and Binary & Count Data Regression

Violation of usual assumptions concerning Normality, Homoscedasticity and Collinearity, Diagnostics using quantile-quantile plots, Logistic and Poisson Regression, Ridge and Lasso Regression. Concepts of GLM.

List of experiments (to be executed using Scientific Calculators and/or R Programming)

Sl. No.	Name of the Experiments
1	Estimability when X is a full rank matrix and not a full rank matrix
2	Simple Linear Regression.
3	Multiple Regression
4	Tests for Linear Hypothesis.
5	Analysis of Variance of a One-way classified data for Fixed Effects Model.
6	Analysis of Variance of a Two-way classified data with one observation per cell for Fixed Effects Model.
7	Analysis of Variance of a Two-way classified data with more than one observation per cell for Fixed Effects Model.
8	Analysis of Variance of a One-way classified data for Random Effects Model.
9	Logistic and Poisson Regression.
10	Ridge and Lasso Regression

Text Books

T1. Sengupta D, Jammalamadaka Rao S. (2003): Linear Models - An Integrated Approach, World Scientific, Series on Multivariate Analysis Vol.6.

T2. Faraway J. J, (2006): Extending the Linear Model with R, Chapman and Hall/CRC, Texts in Statistical Science Series

Reference Books

R1. Scheffe H. (1959): The Analysis of Variance, John Wiley.

R2. Rao C.R, Toutenburg H, Shalabh, Heumann C: Linear Models and Generalizations, 3rd Edition, Springer.

Evaluation:

Mode of Evaluation	Hybrid	
Weightage	Comprehensive and Continuous Assessment	End Semester Examination
	50%	50%

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	2	1	2	1	2	1	2	1	1	2	2	3	2	2	3
CO2	2	2	3	2	2	2	2	1	1	2	3	3	2	2	3
CO3	1	1	2	2	2	2	3	1	1	2	2	1	2	1	3
CO4	2	1	2	3	2	2	2	1	1	2	2	3	2	2	3
CO5	2	1	3	3	2	2	2	1	1	3	2	2	2	2	3
CO6	1	1	2	3	2	3	2	1	1	2	2	1	2	1	3
	1.67	**1.17**	**2.33**	**2.33**	**2.00**	**2.00**	**2.17**	**1.00**	**1.00**	**2.17**	**2.17**	**2.17**	**2.00**	**1.67**	**3.00**

Course Title	**Elementary Data Science**
Course Code	**SDS302**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives:

The primary objective of this course are as follows:

- To make understand the students about bigdata and data extraction techniques.
- To provide a basic concept machine learning algorithm.
- To make student aware about applications of machine learning in real-life problems.
- To provide students a basic concept of recommendation system and its different algorithms.
- To make understand students about data visualization and data privacy.

Course Outcomes:

On completion of this course, the students will be able to

CO1: Explain the data science lifecycle, key components of data science, and differentiate between data science and machine learning.

CO2: Implement and apply basic machine learning algorithms like Linear Regression, k-NN, and k-means for real-world applications such as spam filtering.

CO3: Perform data wrangling, feature generation, and selection using relevant algorithms to prepare data for machine learning models.

CO4: Apply dimensionality reduction techniques like PCA and SVD, and build recommendation systems using clustering and graph partitioning techniques.

CO5: Apply basic principles and tools of data visualization to present and interpret data insights effectively.

CO6: Discuss the ethical, privacy, and security concerns in data science, and understand the role of data scientists in the future.

Course Description:

This course is designed to provide students a fundamental concept of data science. The course deals with data science components along with basics of machine learning algorithms which helps students to understand uses of data science in real-life applications. This course also includes basic concepts of recommendation systems, data visualization and data privacy that can be useful for students to learn pattern of big data and its safety measures. Through this course students may get fundamental ideas of different practical application in industry.

Course Content:

Module I:

Introduction, Big Data and Data Science hype, data science life cycle, data science components, Data science and machine learning, future of machine learning, basic tools, creating data frame, and extracting data from data frame, functions for reading data frame.

Module II:
Three Basic Machine Learning Algorithms, Linear Regression - k-Nearest Neighbors (k-NN), k-means, applications, Filtering Spam, Data Wrangling, Feature Generation and Feature Selection Feature Selection algorithms, Decision Trees, applications in real life

Module III:
Recommendation Systems: Building a User-Facing Data Product, Algorithmic ingredients of a Recommendation Engine - Dimensionality Reduction, Singular Value Decomposition, Principal Component Analysis with applications, clustering of graphs, Partitioning of graphs, related problems.

Module IV:
Data Visualization, Basic principles, ideas and tools for data visualization, data Science and Ethical issues, Discussions on privacy, security, ethics, A look back at Data Science and Next-generation data scientists.

List of Practical's (Use R or Python language platform to perform the following experiments)

1. Creating data frames
2. Extracting data from data frames
3. Functions for reading and writing data
4. Problems based on model fitting
5. Descriptive statistics and graphics
6. Probability and distribution.
7. Problems based on EDA
8. Fitting regression
9. Applications based on decision trees
10. Applications of principal component analysis
11. Application on data clustering
12. Applications on partitioning of graphs
13. Applications based on data visualization techniques

Text Books

T1. Cielen, D., Meysman, A. D. B. & Ali, M. Introducing Data Science: Big Data, Machine Learning, and more, using Python tools. Manning Publications Co., USA.

T2. Reema Thareja, Data Science and Machine Learning with R, Mc Graw Hill, 2021.

T3. Yau, N. (2011). Visualize this: the Flowing Data guide to design, visualization, and statistics. John Wiley & Sons.

Reference Books

R1. Haider, M. Getting Started with Data Science: Making Sense of Data with Analytics. IBM Press.

Evaluation:

Mode of Evaluation	Theory		Practical	
Weightage	Comprehensive and Continuous Assessment	End Semester Examination	Comprehensive and Continuous Assessment	End Semester Examination
	25%	25%	25%	25%

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	2	2	2	2	2	2	2	1	2	2	3	3	2	3
CO2	2	3	3	3	3	3	2	1	1	2	3	3	3	3	3
CO3	2	3	3	3	3	3	2	2	1	2	2	3	3	3	3
CO4	2	2	3	3	3	3	2	2	1	2	2	3	3	3	3
CO5	2	1	2	2	3	2	2	2	1	3	3	2	3	2	3
CO6	2	1	2	3	2	3	3	2	3	2	2	3	3	2	3
	2.17	**2.00**	**2.50**	**2.67**	**2.67**	**2.67**	**2.17**	**1.83**	**1.33**	**2.17**	**2.33**	**2.83**	**3.00**	**2.50**	**3.00**

Course Title	Index Number and Time Series analysis
Course Code	SDS303
Credit	4
Contact Hours (L-T-P)	3-0-2
Course Type	Hybrid

Course Objectives:

The objectives of this course are as follows:

- To understand the construction, uses, and limitations of index numbers in economic analysis.
- To introduce time series data and methods for estimating trends and seasonal components.
- To familiarize students with basic stochastic models, including Moving Average (MA) and Autoregressive (AR) processes, and their applications in analyzing temporal data.

Course Outcomes:

On completion of this course, the students will be able to:

CO1: Explain the construction, uses, and limitations of various index numbers (e.g., Laspeyres', Paasche's, Fisher's, and Chain Index Numbers) and apply them in economic analysis.

CO2: Construct and interpret key index numbers like the Consumer Price Index (CPI), Wholesale Price Index (WPI), and Index of Industrial Production.

CO3: Understand time series components (trend, seasonal, cyclical, random) and apply methods like freehand curve, moving averages, and growth curves for trend estimation.

CO4: Decompose time series data using additive and multiplicative models, and evaluate the effect of eliminating the trend on other components.

CO5: Apply methods like simple averages, ratio-to-moving average, and ratio-to-trend to estimate the seasonal component of time series data.

CO6: Introduce stochastic processes, including stationary time series, Auto-covariance (ACVF), Auto-correlation functions (ACF), and model building using MA and AR processes.

Course Description:

This course introduces index numbers and time series analysis, focusing on their application in economic and industrial data. Students will learn to construct and interpret index numbers, including Consumer Price Index (CPI) and Wholesale Price Index (WPI). The course also covers the decomposition of time series data, methods of trend estimation, and the estimation of seasonal components. Additionally, students will be introduced to stochastic modeling, including Moving Average (MA) and Autoregressive (AR) processes, with an emphasis on analyzing and predicting time-dependent data.

Course Content:

Module I: Index Numbers

Weighted means, Price and Quantity Index Numbers, Value Index, Construction, Uses, Limitations, Laspeyres' and Paasche's Index Numbers. Tests of Index Numbers and Fisher's ideal Index Number, Chain Index Number. Consumer Price Index Number, Wholesale Price Index Number and Index Number of Industrial Production – methods of construction and uses.

Module II: Introduction to Time Series and Estimation of Trend

Introduction to Time Series data, application of Time Series from various fields. Modelling Time Series as deterministic function plus IID errors. Components of a Time Series (Trend, Seasonal and Cyclical patterns, Random error). Decomposition of Time Series. Additive and Multiplicative models. Estimation of Trend: Free hand curve method, method of Moving Averages, fitting various Mathematical Curves and Growth Curves. Effect of elimination of Trend on other components of the Time Series.

Module III: Estimation of Seasonal Component and Introduction to Stochastic Modelling

Estimation of Seasonal component by Method of Simple Averages, Ratio-to-Moving Average, Ratio-to-Trend. Introduction to Stochastic modelling: Concept of Stationarity. Illustration of how a Stationary Time Series may show temporal patterns. Stationarity in mean. Auto-covariance (ACVF) and Auto-correlation functions (ACF) and their properties. Moving-average (MA) process and Autoregressive (AR) process of orders one and two. ACF and its graphical use in guessing the order of MA processes.

List of experiments (to be executed using Scientific Calculators and/or MS Excel/or R)

1. Calculation of Price and Quantity Index Numbers.
2. Applications on Chain Index Numbers.
3. Construction of Consumer and Wholesale Price Index Numbers.
4. Plotting a real-life Time Series and detecting various features (Trend, periodic behaviours etc.)
5. 5 Fitting and plotting of mathematical curves: Linear, Parabolic, Exponential and Modified Exponential
6. Fitting of Trend by Moving Average Method.
7. Measurement of Seasonal indices Ratio-to-Moving Average method.
8. Measurement of Seasonal indices Ratio-to-Trend method.
9. Plotting ACF of a given Time Series.
10. Forecasting by Exponential Smoothing.

Text Books:

T1. Goon A.M., Gupta M.K. & Dasgupta B. (2002): Fundamentals of Statistics, Vol. II, 8 th Edn. The World Press, Kolkata.

T2. Brockwell and Davis (2010): Introduction to Time Series and Forecasting (Springer Texts in Statistics), 2nd Edition

T3. Mukhopadhyay P. (1999): Applied Statistics, Books & Allied Pvt. Ltd.S. K. Mapa, Higher Algebra, Sarat book house.

Reference Books:

R1. Chatfield C. (1980): The Analysis of Time Series – An Introduction, Chapman & Hall.

R2. Kendall M.G. (1976): Time Series, Charles Griffin

R3. Allen R.G.D. (1975): Index Numbers in Theory and Practice, Macmillan

R4. Gupta S. C. & Kapoor V. K. (1975): Fundamentals of Applied Statistics: A Modern Approach. S. Chand & Company

Evaluation:

Mode of Evaluation	Theory		Practical	
Weightage	Comprehensive and Continuous Assessment	End Semester Examination	Comprehensive and Continuous Assessment	End Semester Examination
	25%	25%	25%	25%

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	2	2	2	2	1	2	1	1	2	2	3	3	2	3
CO2	3	2	3	2	2	2	2	1	1	2	3	3	3	3	3
CO3	3	3	3	3	2	2	2	1	1	2	2	3	3	3	3
CO4	3	3	3	3	3	2	2	1	1	2	2	3	3	3	3
CO5	2	2	3	3	3	2	2	1	1	3	3	3	3	2	3
CO6	2	3	3	3	3	3	2	1	2	2	2	3	3	3	3
	2.67	**2.50**	**2.83**	**2.67**	**2.50**	**2.00**	**2.00**	**1.00**	**1.17**	**2.17**	**2.33**	**3.00**	**3.00**	**2.67**	**3.00**

Course Title	**Internship (Summer Project)**
Course Code	**SDS350**
Credit	**4**
Contact Hours (L-T-P)	**- - -**
Course Type	**-**

Course Objectives:
To apply the theory of Statistics and Data Science in relevance to practical solutions

Course Outcomes
On completion of this course, the students will be able to
CO1: Demonstrate the ability to apply academic learning to real-world situations in a professional setting.
CO2: Evaluate personal skills and strengths through practical experience, identifying areas for professional growth.
CO3: Acquire specialized knowledge in specific fields (engineering, marketing, finance, etc.) relevant to the student's career goals.
CO4: Develop a professional network by interacting with industry experts, colleagues, and mentors.
CO5: Proactively research and secure internship opportunities aligned with career aspirations.
CO6: Improve career-related skills such as communication, teamwork, and problem-solving in a professional environment.

Course Description:
Summer internship gives the student an opportunity to bridge theoretical knowledge in practical applications. It is a learning experience that permits students to apply knowledge acquired in the academic classroom within the professional setting. Such experiential learning supplements academic theory, helps the student to identify personal strengths and guides her/him into specialized fields within the profession (Engineering, site works, marketing, media relations, financial management, etc.). Perhaps equally as important is the chance for the student to begin to establish the professional network so essential for access to, and movement within, the profession. The student may personally research internship opportunities and interview for any opportunity that furthers the student’s professional aspirations in the field.

Evaluation:

Mode of Evaluation	Presentation and Report
	End Semester Examination
Weightage (%)	100

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	2	2	2	2	1	2	1	1	2	2	3	3	2	3
CO2	3	2	3	2	2	2	2	1	1	2	3	3	3	3	3
CO3	3	3	3	3	2	2	2	1	1	2	2	3	3	3	3
CO4	3	3	3	3	3	2	2	1	1	2	2	3	3	3	3
CO5	2	2	3	3	3	2	2	1	1	3	3	3	3	2	3
CO6	2	3	3	3	3	3	2	1	2	2	2	3	3	3	3
	2.67	**2.50**	**2.83**	**2.67**	**2.50**	**2.00**	**2.00**	**1.00**	**1.17**	**2.17**	**2.33**	**3.00**	**3.00**	**2.67**	**3.00**

Course Title	**Survey Sampling**
Course Code	**SDS304**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives

- To make students understand the fundamental principles of sample surveys and diverse sampling designs for statistical investigation.
- To enable learners to estimate population parameters and their variances using appropriate sampling techniques.
- To analyze the efficiency, precision, and suitability of different sampling methods under varied practical conditions.
- To develop the ability to compare and select optimal allocation and estimation strategies for real-world survey problems.
- To enable students to design scientifically sound sampling plans and apply advanced estimation methods for reliable decision-making.

Course Outcomes:

On completion of this course, the students will be able to:

CO1: Recall principles of sample surveys, sampling errors, and basic concepts of SRS, stratified, systematic, cluster, two-stage, and randomized response sampling.

CO2: Explain procedures and assumptions of various sampling designs and estimation of population mean, total, and proportion.

CO3: Apply sampling methods to compute estimators, variances, and determine sample size under different designs.

CO4: Analyze and compare efficiency and precision of different sampling techniques and allocation methods.

CO5: Evaluate performance of ratio, regression, Hartley-Ross estimators, and post-stratification in practical survey situations.

CO6: Design an appropriate sampling strategy for real-life survey problems with suitable estimation techniques.

Course Description:

This course introduces the fundamental principles of sample survey, including simple random sampling, estimation of population parameters, and sample size determination. It covers stratified sampling, ratio and regression estimation methods, and advanced allocation techniques for improving precision. The course also explores systematic, cluster, and multi-stage sampling methods with their estimation procedures and variance analysis.

Course Content:

Module I: Simple Random Sampling

Principals of sample survey, Sampling and Non-Sampling Errors, Concept of complete enumeration versus sampling, Types of sampling: non-probability and probability sampling, basic principle of sample survey, simple random sampling with and without replacement, definition and procedure of selecting a sample, estimates of population mean, total and

Course Title	**Machine Learning**
Course Code	**SDS308**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives:

The primary objective of this course is as follows:

- To develop strong foundations in machine learning principles, statistical learning theory, and information-theoretic concepts.
- To build analytical understanding of optimization techniques, metric spaces, and regularization methods for effective model development.
- To enable practical implementation of supervised learning algorithms for accurate classification and informed decision-making.
- To equip learners with unsupervised learning and dimensionality reduction techniques for high-dimensional data analysis.
- To foster the ability to discover meaningful patterns and actionable knowledge using clustering and frequent pattern mining techniques for real-world applications.

Course Outcomes:

On completion of this course, the students will be able to

CO1: Recall key concepts of machine learning, statistical learning theory, information theory, optimization, supervised and unsupervised learning, dimensionality reduction, and pattern mining.

CO2: Explain theoretical foundations of learning models, including statistical principles, information measures, optimization methods, classification, clustering, and association analysis.

CO3: Apply optimization, classification, clustering, dimensionality reduction, and frequent pattern mining techniques to solve machine learning problems.

CO4: Analyze learning models using bias-variance trade-off, loss functions, convergence behavior, performance metrics, and clustering and pattern mining validity measures.

CO5: Evaluate supervised and unsupervised models using regularization, optimization constraints, cross-validation, and information-theoretic and performance metrics.

CO6: Create integrated machine learning solutions by selecting appropriate modeling, optimization, classification, clustering, dimensionality reduction, and pattern mining techniques.

Course Description:

This course introduces the fundamental principles, algorithms, and practical applications of machine learning. Students will explore supervised and unsupervised learning methods, statistical learning theory, optimization techniques, and information theoretic measures. The course emphasizes hands-on implementation of classical machine learning algorithms, dimensionality reduction, clustering, and pattern mining, preparing students for advanced studies in machine learning and data-driven problem solving.

Course Content:

Module I:

Overview of Machine Learning: supervised, unsupervised, and semi-supervised learning; components of ML models; generative vs discriminative modeling paradigms, Empirical risk minimization (ERM), structural risk minimization (SRM), overfitting and underfitting, bias-variance trade-off, regularization principles, cross-validation techniques; Metric Space: concepts of metric space, examples in machine learning, Normed Linear Space.

Module II:

Information Theoretic Measures: entropy, cross-entropy, KL-divergence, mutual information; Feature Engineering: Feature Construction, Sparse Feature Exploration, High-Dimensional Covariance Mapping, Multi-Feature Scatterplots, Feature Ranking; Optimization and Linear Search Method

Module III:

Fundamental concepts of classification model: decision theory, loss functions, Probabilistic classification model, K-NN & Weighted K-NN, Linear Discriminant Analysis, Support Vector Machine, Decision Trees, logistic regression, Classification assessment metrics.

Module IV:

Concepts of representative based clustering, K-Means, Expectation-Maximization algorithm. Agglomerative Hierarchical Clustering. Fundamental concepts of High dimensional data, Data reduction techniques: PCA & SVD

Frequent Pattern Mining: Association analysis with the Apriori algorithm, efficiently finding frequent item sets with FP-growth.

Lab Experiments Using R/Python

1. Gradient Descent and SGD Implementation: Implement gradient descent and stochastic gradient descent on a simple regression problem; analyze convergence, step size, and effect of momentum.
2. K-Nearest Neighbors Classification: Implement K-NN and Weighted K-NN on a real dataset; study effect of K and distance metrics on accuracy.
3. Probabilistic Classification Model: Implement Bayes and Naïve Bayes Algorithm on real/ simulated dataset and study the performance of the model.
4. Support Vector Machine (Linear Case): Implement linear SVM for separable datasets; visualize margin, support vectors, and classification boundaries; compare with logistic regression.
5. Linear Discriminant Analysis: Implement LDA, visualize decision boundaries and evaluate using metrics like accuracy, precision, and recall.
6. Apply K-Means, hierarchical clustering, and Gaussian Mixture Model with EM on real or simulated datasets; visualize clusters and compare clustering performance.
7. Pattern Mining: Apply Apriori and FP-growth algorithms for frequent pattern mining.

8. Dimensionality Reduction; Perform PCA and SVD on high-dimensional data, analyze dimensionality reduction effects on clustering or association tasks

Text Books:

T1. Kevin P. Murphy (2022), Probabilistic Machine Learning: An Introduction, MIT Press.

T2. Mohammed J. Zaki, Wagner Meira, JR (2020), Data Mining and Machine Learning: Fundamental Concepts and Algorithm, Cambridge University Press, 2nd Edition

Reference Books:

R1. Trevor Hastie, Robert Tibshirani, Jerome Friedman (2009), The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Edition, Springer (Springer Series in Statistics)

R2. Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar (2018), Foundations of Machine Learning, 2nd Edition, MIT Press.

Evaluation:

Mode of Evaluation	Hybrid			
Weightage	Comprehensive and Continuous Assessment		End Semester Examination	
	Theory	Practical	Theory	Practical
	25%	25%	25%	25%

proportion, variances of these estimates, estimates of their variances and sample size determination.

Module II: Stratified Random Sampling
Stratified random sampling: Technique, estimates of population mean and total, variances of these estimates, proportional and optimum allocations, and their comparison with SRS. Practical difficulties in allocation, estimation of gain in precision, post stratification and its performance.

Module III: Systematic, Cluster and Two-Stage Sampling Methods
Systematic Sampling: Technique, estimates of population mean and total, variances of these estimates $(N = n \times k)$. Cluster sampling (equal clusters only) estimation of population mean and its variance, comparison (with and without randomly formed clusters). Two-stage Sampling, Estimation of Population mean and variance of the estimate. Introduction to Ratio and regression methods of estimation

List of Practical to be performed using R/Python:

1. Select an SRS with and without replacement.
2. For a population of size 5, estimate population mean, population mean square and population variance. Enumerate all possible samples of size 2 by WR and WOR and establish all properties relative to SRS.
3. For SRSWOR, estimate mean, standard error, the sample size
4. Stratified Sampling: allocation of sample to strata by proportional and Neyman's methods Compare the efficiencies of the above two methods relative to SRS
5. Estimation of gain in precision in stratified sampling.
6. Comparison of systematic sampling with stratified sampling and SRS in the presence of a linear trend.
7. Ratio and Regression estimation: Calculate the population mean or total of the population. Calculate mean squares. Compare the efficiencies of ratio and regression estimators relative to SRS.
8. Cluster sampling: estimation of mean or total, variance of the estimate, an estimate of the intra-class correlation coefficient, and efficiency as compared to SRS.

Text Books:

T1. Parimal Mukhopadhyay (2015): Theory and Methods of Survey Sampling, PHI Learning Private Limited, New Delhi, 2nd Edition.

T2. Arijit Choudhuri (2010): Essentials of Survey Sampling, PHI Learning Private Limited, New Delhi, 2nd Edition.

Reference Books:

R1. William G. Cochran (2015): Sampling Techniques, Wiley Publication, 3rd Edition

R2. M.N. Murthy (1977): Sampling Theory & Statistical Methods, Statistical Pub. Society, Calcutta.

Evaluation:

Mode of Evaluation	Hybrid	
Weightage	Comprehensive and Continuous Assessment	End Semester Examination
	50%	50%

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	2	3	2	2	2	1	1	1	2	2	3	3	2	3
CO2	2	3	2	3	2	2	2	1	1	2	3	2	3	2	2
CO3	2	2	3	2	3	3	2	1	1	2	2	3	3	3	3
CO4	3	1	2	2	1	2	3	1	1	2	3	2	3	3	3
CO5	2	2	3	3	3	2	2	1	1	3	2	2	3	2	3
CO6	2	2	3	3	2	3	2	2	1	2	3	3	3	3	3
	2.33	**2.00**	**2.67**	**2.50**	**2.17**	**2.33**	**2.00**	**1.17**	**1.00**	**2.17**	**2.50**	**2.50**	**3.00**	**2.50**	**2.83**

Course Title	**Applied machine learning**
Course Code	**SDS305**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives:

- To develop students' ability to build machine learning algorithms for real-life challenges.
- To understand the different supervised and unsupervised machine learning algorithms and its applications.
- To enable students to understand, interpret, and process raw data

Course Outcomes:

On the completion of this course the student will be able to

CO1: Explain key machine learning concepts, including problem formulation, forms of learning (supervised, unsupervised, reinforcement), and learning inspired by natural processes.

CO2: Implement classifiers like k-Nearest Neighbor, Naive Bayes, SVM, Decision Trees, and ensemble methods and evaluate their performance using relevant metrics.

CO3: Apply regression techniques such as multiple linear regression, logistic regression, and tree-based regression, while understanding concepts like bias/variance tradeoff and error metrics.

CO4: Understand learning theory concepts, including the Vapnik–Chervonenkis (VC) dimension, union and Chernoff/Hoeffding bounds, and apply them to assess model performance.

CO5: Implement unsupervised learning techniques like clustering (k-means, hierarchical clustering) and association analysis for grouping and finding patterns in unlabelled data.

CO6: Utilize advanced machine learning models, including support vector machines (SVM) and Bayesian belief networks, for classification and decision-making tasks.

Course Description:

This course is aimed to develop a student's ability to build machine learning algorithm for real life challenges. Students will learn about different supervised and unsupervised machine learning algorithms. They will also be able to understand, interpret, and process raw data and perform feature extraction techniques.

Course Content:

Module I:

Overview of Machine Learning, Well-Posed Machine Learning Problems, Forms of Learning, Supervised/Directed Learning, Unsupervised/Undirected Learning, Reinforcement Learning, Learning Based on Natural Processes: Evolution, Swarming, and Immune Systems.

Module II:

Classifying with k-Nearest Neighbor classifier, Naive Bayes Classifier, Bayesian Belief Networks, Bayes Decision Theory, Linear Discriminant Functions (SVM), Support vector machine classifier, Decision Tree classifier, Bagging, Boosting, Random Forest, performance evaluation metrics.

Module III:

Forecasting and Learning Theory: Predicting numeric values: regression, Multiple linear regression, Logistic regression, Expectation and Minimization, Tree-based regression, Bias/variance tradeoff, Union and Chernoff/Hoeffding bounds, Vapnik– Chervonenkis (VC) dimension, Error metrics.

Module IV:

Unsupervised Learning, Clustering, grouping unlabeled items using k-means clustering, hierarchical clustering, Association analysis with the Apriori algorithm, efficiently finding frequent item sets with FP-growth.

Lab Experiments:

Module -I

1. Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.
2. Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.
3. Assuming a set of documents that need to be classified, use the naïve Bayesian Classifier model to perform this task. Calculate the accuracy, precision, and recall for your data set.
4. Apply EM algorithm to cluster a set of data stored in a .CSV file. Use the same data set for clustering using k-Means algorithm. Compare the results of these two algorithms and comment on the quality of clustering. You can add Java/Python ML library classes/API in the program.
5. Write a program to implement K-NN algorithm to classify a real data set. Print both correct and wrong predictions. Java/Python ML library classes can be used for this problem.
6. Implement the linear Regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs.
7. Implement the Multiple linear regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs.

Module II

Implementation through mini project using ML Concept

Text Books:

T1. Machine Learning – Tom M. Mitchell

T2. Introduction to Machine Learning - Nils J Nilsson

T3. Introduction to machine learning with Python: a guide for data scientists. - Muller, A. C., & Guido, S. (2017), O'Reilly Media.

Reference Books:

R1. Pattern Recognition and Machine Learning – Christopher M. Bishop

R2. Applied Machine learning - Dr. M. Gopal, McGraw Hill Education (India) Private Limited

R3. An Introduction to Statistical Learning with Applications in R,- Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Springer New York, NY, 2013.

Evaluation:

Mode of Evaluation	Hybrid			
Weightage	Comprehensive and Continuous Assessment		End Semester Examination	
	Theory	Practical	Theory	Practical
	25%	25%	25%	25%

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	2	2	2	2	1	2	1	1	2	2	3	3	2	3
CO2	3	3	3	2	3	2	2	1	1	2	2	3	3	3	3
CO3	3	3	3	3	3	2	2	1	1	2	2	3	3	3	3
CO4	3	3	3	3	3	2	2	1	1	2	2	3	3	3	3
CO5	3	3	3	3	3	2	2	1	1	3	2	3	3	3	3
CO6	3	3	3	3	3	3	2	1	1	2	2	3	3	3	3
	3.00	**2.83**	**2.83**	**2.67**	**2.83**	**2.00**	**2.00**	**1.00**	**1.00**	**2.17**	**2.00**	**3.00**	**3.00**	**2.83**	**3.00**

Course Title	**Database Management System**
Course Code	**SDS306**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives:

- To understand database concepts, applications, data models, schemas and instances.
- To implement the relational database design and data modelling using entity-relationship (ER) model.
- To demonstrate the use of constraints and relational algebra operations.
- To be able to use SQL in querying the database.
- To demonstrate Normalization process.
- To learn the new emerging Technologies and Applications in database.

Course Outcomes:

On completion of this course, the students will be able to

CO1: Define DBMS, its characteristics, advantages, and applications, and understand key concepts like schemas, instances, and data independence.

CO2: Compare different data models, and understand the Entity-Relationship (E-R) model, including entities, attributes, keys, and how to design E-R diagrams.

CO3: Use relational algebra operations (selection, projection, joins) and understand functional dependencies, database anomalies, and normalization techniques (1NF to 5NF).

CO4: Explain the concepts of Relational Database Management Systems (RDBMS), its components, and apply Codd's rules in relational database design.

CO5: Use SQL to create tables, add constraints, and manipulate data with DDL, DML, and DCL commands, and apply various forms of SELECT queries with operators.

CO6: Create SQL functions for data aggregation, join multiple tables (including self-joins), and understand transaction control and other SQL command functionalities.

Course Description:

Databases form the backbone of all applications today – tightly or loosely coupled, intranet or internet based, financial, social, administrative, and so on. Database Management Systems (DBMS) based on relational and other models have long formed the basis for such databases. Consequently, Oracle, Microsoft SQL Server, Sybase etc. have emerged as leading commercial systems while MySQL, PostgreSQL etc. lead in open source and free domain.

While DBMS's differ in the details, they share a common set of models, design paradigms and a Structured Query Language (SQL). In this background the course examines data structures, file organizations, concepts and principles of DBMS's, data analysis, database design, data modeling, database management, data & query optimization, and database implementation. More specifically, the course introduces relational data models; entity-relationship modeling, SQL, data normalization, and database design. Further it introduces query coding practices using MySQL (or any other open system) through various assignments. Design of simple multi-tier client / server architectures based and Web-based database applications is also introduced.

Course Content:

Module I

DBMS Definition, Characteristics of DBMS, Application and advantages of DBMS, Instances, Schemas and Database States, Three Levels of Architecture, Data Independence, DBMS languages, Data Dictionary, Database Users, Data Administrators.

Module II

Data Models, types and their comparison, Entity Relationship Model, Entity Types, Entity Sets, Attributes and its types, Keys, E-R Diagram, Data Integrity RDBMS –Concept, Components and Codd's rules.

Module III

Relational Algebra (selection, projection, union, intersection, Cartesian product, Different types of join like theta join, equi-join, natural join, outer join) Functional Dependencies, Good & Bad Decomposition, Anomalies as a database: A consequences of bad design, Normalization: 1NF, 2NF, 3NF, BCNF, 4NF 5NF.

Module IV

Introduction to SQL, DDL, DML, and DCL statements, Creating Tables, Adding Constraints, Altering Tables, Update, Insert, Delete & various Form of SELECT- Simple, Using Special Operators for Data Access. Aggregate functions, Joining Multiple Tables (Equi Joins), Joining a Table to itself (self Joins) Functions. Structured Query Language (SQL), Data Definition Language Commands, Data Manipulation Language Commands, Transaction Control Commands, SQL Command.

Lab Topics:

Experiment 1: Familiarization of structured query language.
Experiment 2: Table Creation.
Experiment 3: Insertion, Updation, Deletion of tuples.
Experiment 4: Executing different queries based on different functions.
Experiment 5: Performing joining operations.
Experiment 6: Nested Queries.
Experiment 7: Use of aggregate functions.
Experiment 8: Use of group functions.
Experiment 9: Use of order by functions.
Experiment 10: Arithmetic operations.
Experiment 11: Trigger using SQL.
Experiment 12: Introduction to PL/SQL.
Experiment 13: Report generation of various queries.
Experiment 14: Merging Data Bases with front end using ODBC connection.
Experiment 15: SQL Injection on a non-harmful test page.

Text Books

T1. Elmasri, R., & Navathe, S. (2010). Fundamentals of database systems. Addison-Wesley Publishing Company.

T2. Silberschatz, A., Korth, H. F. & Sudarshan, S. Database System Concepts, 6th Edition. McGraw Hill.

T3. Date, C. (2012). Database Design and Relational Theory: Normal Forms and All That Jazz. O'Reilly Media, Inc.

Reference Books

R1. Date, C. J. (2011). SQL and relational theory: How to write accurate SQL code. O'Reilly Media, Inc

Evaluation:

Mode of Evaluation	Theory		Practical	
Weightage	Comprehensive and Continuous Assessment	End Semester Examination	Comprehensive and Continuous Assessment	End Semester Examination
	25%	25%	25%	25%

Mapping COs with POs & PSOs:

CO/PO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**	**PSO1**	**PSO2**	**PSO3**
CO1	3	2	2	2	3	3	1	1	1	2	2	3	3	2	3
CO2	3	2	3	3	3	3	2	1	1	2	2	3	3	3	3
CO3	3	2	3	3	3	3	2	1	1	2	2	3	3	3	3
CO4	3	3	3	3	2	3	1	1	1	2	2	3	3	3	3
CO5	3	2	3	3	3	3	2	1	1	2	2	3	3	3	3
CO6	3	2	3	3	3	3	2	1	1	2	2	3	3	3	3
	3.00	**2.17**	**2.83**	**2.83**	**2.83**	**3.00**	**1.67**	**1.00**	**1.00**	**2.00**	**2.00**	**3.00**	**3.00**	**2.83**	**3.00**

Course Title	**Quantitative Methods for Finance**
Course Code	**SDS307**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives:

- To understand introductory knowledge of financial risk through mathematical applications and learn the basics of financial analysis in terms mathematics and statistics.
- To solve and analysis some basic problems in financial risk management, portfolio optimization and asset management by using their mathematical and statistical concept and skill.

Course Outcomes:

After completion of the course the students will be able to

CO1: Define inverse functions, exponential and logarithmic functions, and Taylor expansions in financial contexts, with a focus on risk factors and sensitivities.

CO2: Interpret portfolio holdings, portfolio weights, and returns (percentage, log, discrete, and continuous compounding) in finance.

CO3: Determine eigenvalues and eigenvectors in portfolio risk and return analysis, and understand the positive definiteness of covariance and correlation matrices.

CO4: Apply univariate probability distributions (binomial, Poisson, normal) in finance and use linear regression (OLS) for hypothesis testing and goodness of fit.

CO5: Use numerical methods like bisection, Newton-Raphson, and gradient methods to calculate financial risk and bond yields, and apply interpolation techniques in financial applications.

CO6: Discuss utility theory, risk preferences, mean-variance optimization, and portfolio diversification to solve problems related to portfolio allocation and efficient frontiers.

Course Description:

This course will enable students to apply mathematical knowledge into financial domain. They can implement several statistical and numerical techniques as well in the financial problems. Several terminologies in the context of financial models will be made familiar to the students and eventually they will be able to apply techniques from linear algebra to solve problems in financial realm.

Course Content:

Module I

Basic Calculus for Finance: Functions & inverse function with example in finance, exponential and natural logarithm function with example in finance, monotonic, concave and convex functions, Taylor expansion, Risk Factors and their Sensitivities.

Analysis of financial returns: Portfolio holdings and portfolio weights, profit and loss, percentage and log returns, discrete and continuous compounding in discrete time, return on a linear portfolio and its applications in finance.

Module II:

Quadratic Forms, eigenvalues and eigenvectors of a 2×2 correlation matrix, eigenvalue test for definiteness.

Applications to linear portfolios: Portfolio risk and return, positive definiteness of covariance and correlation Matrices.

Module III

Review of univariate probability distributions, binomial distribution, poisson and exponential distributions, normal distribution, application in finance, introduction to linear regression, Ordinary Least Squares, properties of the error process, ANOVA and goodness of fit, hypothesis tests on coefficients.

Module IV

Numerical Methods in Finance: Calculation of risk and bond yield using bisection, Newton-Raphson method and gradient methods;

Interpolation and extrapolation: Polynomial and Cubic Interpolation, some applications in finance.

Module V

Introduction to portfolio theory, Utility Theory and its properties, Risk preference, Coefficients of Risk Aversion, some standard utility functions, mean–variance criterion, portfolio allocation, portfolio diversification, minimum variance portfolios, the Markowitz problem, efficient frontier, optimal allocation.

Practical to be performed using Excel/R/MATLAB packages

1. Find the stationary points of a function and test the convexity and concavity
2. Find the returns of a Linear Portfolio
3. Finding Eigenvalues to test the Definiteness of a Covariance and Correlation
4. Finding Eigenvalues and Eigenvectors of Covariance and Correlation Matrices for linear portfolio
5. Perform Error Analysis, ANOVA and goodness of fit, Hypothesis Tests on Coefficients for a linear regression problem
6. Applications of Linear Regression in Finance
   - Testing a Theory
   - Analysing Empirical Market behaviour
   - Optimal Portfolio Allocation
7. Finding root of an equation using Bisection Method
8. Finding root of an equation using Newton–Raphson Iteration
9. Finding minimum and maximum using Gradient Methods
10. Finding Interpolating polynomial using Polynomial Interpolation for Currency Options
11. Finding the allocation and volatility of a Minimum Variance Portfolios of financial data sets
12. Solve the Markowitz problem in finance.

Text Books

T1. Alexander, C. (2008). Quantitative methods in finance. Wiley.

T2. Skoglund, J., & Chen, W. (2015). Financial risk management: Applications in market, credit, asset and liability management and firmwide risk. John Wiley & Sons.

Evaluation:

Mode of Evaluation	Hybrid			
Weightage	Comprehensive and Continuous Assessment		End Semester Examination	
	Theory	Practical	Theory	Practical
	25%	25%	25%	25%

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	2	2	2	3	3	1	1	1	2	2	3	3	2	3
CO2	3	2	3	3	3	3	2	1	1	2	2	3	3	3	3
CO3	3	2	3	3	3	3	2	1	1	2	2	3	3	3	3
CO4	3	3	3	3	2	3	1	1	1	2	2	3	3	3	3
CO5	3	2	3	3	3	3	2	1	1	2	2	3	3	3	3
CO6	3	2	3	3	3	3	2	1	1	2	2	3	3	3	3
	3.00	**2.17**	**2.83**	**2.83**	**2.83**	**3.00**	**1.67**	**1.00**	**1.00**	**2.00**	**2.00**	**3.00**	**3.00**	**2.83**	**3.00**

Course Title	**Project Work**
Course Code	**SDS351**
Credit	**4**
Contact Hours (L-T-P)	**- - -**
Course Type	

Course Objectives:

1. To address the real-world problems and find the required solution.
2. To fabricate and implement the mini project intended solution for project-based learning
3. To improve the team building, communication and management skills of the students

Course Outcomes

On completion of this course, the students will be able to

CO1: Use statistical methods to solve practical problems in various fields such as data science, agriculture, finance, and industry.

CO2: Analyze statistical concepts through hands-on application in project work.

CO3: Interpret data, drawing valid conclusions from statistical results.

CO4: Design and conducting projects, enhancing their ability to perform independent research and critical analysis.

CO5: build a foundation for undertaking larger, more complex projects in their future careers by applying theoretical knowledge to practical tasks.

CO6: Develop relevant skills and experiences to enhance their resumes and improve their competitiveness in the job market.

Course Description:

The role of project dissertation in Statistics and Data Science is very important. Project helps a student to explore and strengthen the understanding of fundamentals through practical application of theoretical concepts. Statistics is the basis of many areas including data science, agriculture, finance, industry. Everything around us needs statistical analysis for acceptance. Project allows a student to apply Statistical theory in practical purpose to draw some valid conclusions. It acts like a beginner's guide to do larger projects later in their career. It not just affects the grades of the program but also matter a lot for good CV/Resume. So before choosing the minor and major project, you should explore the options and pick the correct domain where the opportunities are immense.

Evaluation:

Mode of Evaluation	Presentation and Report
	End Semester Examination
Weightage (%)	100

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	2	2	1	3	3	2	1	1	2	2	3	3	3	3
CO2	3	2	3	2	3	3	2	1	1	2	2	3	3	3	3
CO3	3	3	3	2	3	3	2	1	1	2	2	3	3	3	3
CO4	3	3	3	2	3	3	2	1	1	2	2	3	3	3	3
CO5	3	2	3	3	3	3	3	1	1	2	2	3	3	3	3
CO6	3	2	3	3	3	3	3	1	1	2	2	3	3	3	3
	3.00	**2.33**	**2.83**	**2.17**	**3.00**	**3.00**	**2.33**	**1.00**	**1.00**	**2.00**	**2.00**	**3.00**	**3.00**	**3.00**	**3.00**

Course Title	**Stochastic Processes & Queueing Theory**
Course Code	**SDS401**
Credit	**4**
Contact Hours (L-T-P)	**3-1-0**
Course Type	**Theory**

Course Objectives:

The objectives of this course are as follows:

- To develop a strong understanding of probability theory, including probability generating functions and stochastic processes.
- To make students understand both discrete-time and continuous-time Markov chains, focusing on state classification, transition matrices, and long-term behavior.
- To enable students to apply queueing models (e.g., M/M/1) to analyze performance metrics in various real-world scenarios like customer service and telecommunications.
- To enhance analytical and critical thinking skills through problem-solving and case studies, recognizing applications across fields such as operations research and risk management.

Course Outcomes:

On completion of this course, the students will be able to:

CO1: Recall and describe fundamental concepts of probability generating functions and stationary processes.

CO2: Explain the principles of discrete-time Markov chains, including state classification and Chapman-Kolmogorov equations.

CO3: Apply Markov chain models to solve queuing problems, including the M/M/1 model and simple Markovian queuing.

CO4: Analyze continuous-time Markov chains and interpret their applications in queueing networks, communication systems, and stochastic processes.

CO5: Evaluate the gambler's ruin problem and determine the expected duration of the game using stochastic modeling.

CO6: Design and implement solutions for real-world stochastic problems using queuing models, communication systems, and Markov processes.

Course Description:

This course introduces stochastic processes and Markov chains, focusing on probability generating functions, discrete-time Markov chains, and continuous-time Markov chains. Students will explore key concepts such as transition probability matrices, classification of states, and applications in queuing models and communication systems. The course combines theoretical foundations with practical applications, equipping students with essential skills for analyzing stochastic systems.

Course Content:

Module I: Probability

Probability generating function and its properties, Bivariate probability generating function, Sequence of random variables; Simple stochastic process: Introduction, Stationary Process

Module II: Markov chain and Markov process

Discrete-time Markov chains: Introduction, Definition and Transition Probability Matrix, Chapman-Kolmogorov Equations, Classification of States, Limiting and Stationary Distributions, Ergodicity, Time Reversible Markov Chain, Application of Irreducible Markov chains in Queuing Models; Reducible Markov Chains

Module III: Queueing Process

Continuous-time Markov chains: Definition, Kolmogorov Differential Equation and Infinitesimal Generator Matrix, Limiting and Stationary Distributions, Birth and Death Processes, Poisson processes, M/M/1 Queuing model; Simple Markovian Queuing, Applications of Continuous-time Markov chain, Queueing networks, Communication systems, Stochastic Petri Nets, Gambler's Ruin Problem: Classical ruin problem, expected duration of the game.

.

Text Books:

- **T1.** J. Medhi, Stochastic Processes, New Age International (P) Ltd., New Delhi,2nd Edition, 2001
- **T2.** S.M. Ross, Stochastic Processes, 2nd Edition, Wiley, 1996 (WSE Edition).

Reference Books:

- **R1.** H.M. Taylor and S. Karlin, An Introduction to Stochastic Modeling, 3rd Edition, Academic Press, New York, 1998.
- **R2.** Liliana Blanco Castaneda, Viswanathan Arunachalam, Selvamuthu Dharmaraja, Introduction to Probability and Stochastic Processes with Applications, Wiley, 2012.
- **R3.** G. R. Grimmett and D. R. Stirzaker, Probability and Random Processes, 3rd Edition, Oxford University Press, 2001

Evaluation:

<table>
<tr><td>Mode of Evaluation</td><td colspan="2">Te</td></tr>
<tr><td rowspan="2">Weightage</td><td>Comprehensive and Continuous Assessment</td><td>End Semester Examination</td></tr>
<tr><td>50%</td><td>50%</td></tr>
</table>

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	3	1	1	1	1	1	1	1	1	1	2	3	2	3
CO2	3	3	2	2	2	1	1	1	1	2	2	2	3	2	3
CO3	3	3	3	3	2	2	1	1	1	2	2	2	3	3	3
CO4	3	3	3	3	3	2	2	2	1	3	2	3	3	3	3
CO5	3	3	3	3	3	3	2	2	2	3	2	3	3	3	3
CO6	3	3	3	3	3	3	3	3	2	3	3	3	3	3	3
	3	**3**	**2.5**	**2.5**	**2.33**	**2**	**1.67**	**1.67**	**1.33**	**2.33**	**2**	**2.5**	**3**	**2.67**	**3**

Course Title	**Advanced Machine Learning and Applications**
Course Code	**SDS402**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives:

- Develop a strong understanding of advanced machine learning algorithms.
- Enable critical evaluation of algorithm strengths, limitations, and applications.
- Provide hands-on experience with public machine learning toolboxes.
- Foster skills in designing and conducting experiments for model evaluation.
- Cultivate the ability to analyze and interpret experimental results effectively.

Course Outcomes:

On completion of this course, the students will be able to:

CO1: Define and describe the key properties of kernel methods, including kernels for structured data and generative models.

CO2: Explain the concepts of hard and soft margin SVM, PAC theory, and their applications in supervised learning.

CO3: Implement and apply Adaboost and Gradient Boosting algorithms to solve binary and multiclass classification problems.

CO4: Analyze graphical models for structured prediction, including MAP inference, sampling methods, and variational inference techniques.

CO5: Evaluate the effectiveness of supervised and unsupervised dictionary learning methods in tasks like image reconstruction.

CO6: Design and develop a Deep Reinforcement Learning agent using Q-learning to solve real-world decision-making problems.

Course Description:

This course provides a comprehensive exploration of advanced machine learning techniques, including deep learning, distance metric learning, and domain adaptation. Students will gain the ability to explain the functioning of these algorithms and critically analyze their advantages, limitations, and potential applications in various fields. The course also emphasizes practical skills, offering hands-on experience with public domain machine learning toolboxes to solve real-world problems. Additionally, students will learn to design and conduct experiments to evaluate machine learning models and assess their performance through rigorous analysis.

Course Content:

Module I

Kernel Machines: Kernel properties, Kernels for structure data and text, Multiple kernel learning, Generative models

Module II

Variants of Support Vector Machine: Hard and soft margin SVM, Online SVM, Distributed SVM, PAC Theory

Module III

Boosting: Adaboost, Gradient boosting, Structured Prediction and Graphical Models: Learning directed and undirected models, Sampling, MAP inference and prediction, variational inference, causality

Module IV

Dictionary Learning: Fundamentals, Regularization, Supervised and unsupervised dictionary, learning, Transform Learning, Deep Reinforcement Learning.

List of practical to be performed using Python:

1. Implementation of kernel methods for structured data and applying them to real-world datasets.
2. Implementation of Generative models using kernel methods.
3. Implementing hard and soft margin SVM from scratch and using libraries (e.g., scikit-learn) for comparison.
4. Testing PAC theory using SVM on classification tasks.
5. Implementing Adaboost and Gradient Boosting algorithms on various datasets (binary and multiclass).
6. Structured prediction using graphical models: Learning directed models and performing inference using sampling methods.
7. Implementing supervised and unsupervised dictionary learning for image reconstruction tasks.
8. Implementing Transform learning for classification tasks.
9. Building a simple Deep Reinforcement Learning agent using Q-learning.

Text Books

T1. N. CRISTIANINI, J. S-TAYLOR (2000), An Introduction to Support Vector Machines and Other Kernel- based Learning Methods, Cambridge University Press, 1st Edition.

T2. B. SCHOLKOPF, A. J. SMOLA (2001), Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, The MIT Press, 2001, 1st Edition.

Reference Books

R1. R. S. SUTTON, A. G. BARTO (2018), Reinforcement Learning: An Introduction, The MIT Press, 2nd Edition.

R2. D. KOLLER, N. FRIEDMAN (2009), Probabilistic Graphical Models: Principles and Techniques, MIT Press.

Evaluation:

Mode of Evaluation	Hybrid			
Weightage	Comprehensive and Continuous Assessment		End Semester Examination	
	Theory	Practical	Theory	Practical
	25%	25%	25%	25%

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	3	2	1	1	1	1	1	1	1	1	2	3	2	3
CO2	3	3	3	2	2	2	1	1	1	2	2	3	3	2	3
CO3	3	3	3	3	3	2	1	1	1	3	2	3	3	3	3
CO4	3	3	3	3	3	3	2	2	1	3	3	3	3	3	3
CO5	3	3	3	3	3	3	2	2	2	3	3	3	3	3	3
CO6	3	3	3	3	3	3	3	3	2	3	3	3	3	3	3
	3	**3**	**2.83**	**2.5**	**2.5**	**2.33**	**1.67**	**1.67**	**1.33**	**2.5**	**2.33**	**2.83**	**3**	**2.67**	**3**

Course Title	**Pricing, Hedging and Optimization**
Course Code	**SDS403**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives:

- To provide students with a comprehensive understanding of financial markets, focusing on bonds, interest rates, futures, and swaps.
- To learn the theoretical relationships between spot, forward, and futures prices, as well as develop expertise in option pricing models, including the Binomial and Black-Scholes-Merton models.
- To introduce key concepts of asset pricing, including the Capital Asset Pricing Model (CAPM) and portfolio theory, and to develop skills in portfolio optimization and evaluation.

Course Outcomes:

On completion of this course, the students will be able to:
CO1: Define the fundamentals of financial markets and instruments.
CO2: Explain the Binomial and Black-Scholes models for options pricing.
CO3: Develop strategies to optimize portfolios and evaluate asset pricing models.
CO4: Examine classical strategies for speculation and risk management using options.
CO5: Evaluate replicating portfolios for option pricing and trading strategies.
CO6: Discuss different measures for portfolios and adjust for volatility.

Course Description:

This course offers an in-depth understanding of financial markets, focusing on the characteristics of options, bonds, and interest rates, as well as the relationships between spot, forward, and futures prices. Students will explore and analyze various option pricing models, including the Binomial and Black-Scholes-Merton models, while gaining insights into hedging strategies and volatility patterns. The course also covers key asset pricing theories, such as the Capital Asset Pricing Model (CAPM), emphasizing the role of beta in portfolio management. Additionally, students will learn to evaluate portfolio performance using a range of performance measurement tools.

Course Content:

Module I

Introduction, financial & futures markets, bonds and swaps, interest rates, characteristics of bonds and Interest Rates, duration and convexity; futures and forwards, theoretical relationships between spot, forward, and futures.

Module II

Introduction to Option, speculation with Options and classical strategies, the single-period Binomial model, the multi-period Binomial model, replication portfolio for European Options, risk-neutral valuation, hedging Options, trading Options, the Black–Scholes–Merton model and its Greeks; volatility, implied volatility, volatility smiles and skews.

Module III

Theory of asset pricing: capital market theory, capital assets pricing model- the capital market line, beta of an asset, beta of a portfolio, security market line, index tracking optimization models, portfolio performance evaluation measures, replicating portfolios.

Practical to be performed using R/Python/MATLAB packages or through Project work:

1. Develop your strategy to create different payoff combining the options.
2. Write a program for the Binomial Model to calculate different sensitivities by calculating the option prices with different parameters.
3. Write a program to obtain Black-Scholes prices.
4. Write a program to see the changes of BS prices with maturities and strikes.
5. Write a program to obtain Black-Scholes prices from the BINOMIAL model
6. Write a program to obtain different Greeks from Black-Scholes model.
7. Write a program to estimate IV and others volatility.
8. Write a program to find final value of a replicating portfolio and final value of a given option.
9. Write a program to calculate the Sharpe ratios of both the fund and the benchmark and the adjustments for autocorrelation and higher moments.

Text Books

T1. Alexander, C. (2008): Market Risk Analysis Volume III: Pricing, Hedging and Trading Financial Instruments, John Wiley & Sons Ltd.

T2. Alòs, E., Merino, R. (2023): Introduction to Financial Derivatives with Python, CRC Press.

T3. John Hull. Options, Futures, and Other Derivatives. 2009. Fifth Edition. Prentice Hall.

T4. Alexander, C. (2008). Quantitative methods in finance. Wiley.

T5. Skoglund, J., & Chen, W. (2015). Financial risk management: Applications in market, credit, asset and liability management and firmwide risk. John Wiley & Sons.

Evaluation:

Mode of Evaluation	Hybrid			
Weightage	Comprehensive and Continuous Assessment		End Semester Examination	
	Theory	Practical	Theory	Practical
	25%	25%	25%	25%

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	2	1	2	1	2	1	1	1	2	2	1	3	2	2
CO2	2	3	2	2	3	3	2	1	1	3	2	2	2	3	3
CO3	2	2	3	3	3	3	2	2	1	3	3	2	3	2	3
CO4	3	3	2	3	3	3	2	2	2	3	3	2	3	3	3
CO5	3	2	3	3	3	3	2	1	2	3	3	2	3	3	3
CO6	3	2	3	3	3	3	3	3	2	3	3	3	3	3	3
	2.67	**2.33**	**2.33**	**2.67**	**2.67**	**2.83**	**2.00**	**1.67**	**1.50**	**2.83**	**2.67**	**2.00**	**2.83**	**2.67**	**2.83**

Course Title	**Design of Experiments**
Course Code	**SDS404**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives:

The objectives of this course are as follows:

- Explore the fundamental concepts and historical context of experimental designs, including treatments, blocks, and experimental error.
- Learn to implement and analyze Completely Randomized, Randomized Block, and Latin Square Designs, including their layouts and statistical evaluations.
- Gain insights into the advantages and analysis of factorial experiments, including the design of 2^n factorials and the concept of confounding.
- Investigate Balanced Incomplete Block Designs (BIBD), focusing on their parameters, properties, and various types, including symmetric and resolvable designs.

Course Outcomes:

On completion of this course, the students will be able to:

CO1: Define the basic principles, terminology, and concepts of experimental designs, including treatments, blocks, and experimental errors.

CO2: Explain the layout, models, and statistical analysis of CRD, RBD, and LSD, including the analysis with missing observations.

CO3: Make use of statistical analysis of factorial experiments, including complete and partial confounding in 2n factorials.

CO4: Analyze incomplete block designs such as BIBD, resolvable BIBD, and residual BIBD using intra-block analysis techniques.

CO5: Evaluate the efficiency of various experimental designs and determine the relationships among parameters in BIBD.

CO6: Design and execute experiments using scientific calculators or software to analyze CRD, RBD, LSD, factorial experiments, and BIBD, demonstrating a practical understanding of experimental design.

Course Description:

This course provides a comprehensive study of experimental design in agriculture, covering fundamental concepts, terminology, and historical perspectives. Students will explore various design methodologies, including completely randomized designs, factorial experiments, and incomplete block designs, with a focus on statistical analysis and practical applications in agricultural research.

Course Content:

Module I: Experimental designs

Role, historical perspective, terminology: Treatments, Experimental units & Blocks, Experimental error, Basic principles of Design of Experiments (Fisher). Uniformity trials, fertility contour maps, choice of size and shape of plots and blocks in Agricultural experiments.

Basic designs: Completely Randomized Design (CRD), Randomized Block Design (RBD), Latin Square Design (LSD) – layout, model and statistical analysis, relative efficiency, analysis with missing observations.

Module II: Factorial Experiments

Advantages, notations, and concepts, 2^2, $2^3 \ldots 2^n$ Factorial experiments, design and analysis, Total and Partial confounding for 2^n $(n \leq 5)$. Factorial experiments in a single replicate.

Module III: Incomplete Block Designs

Balanced Incomplete Block Design (BIBD) – parameters, relationships among its parameters, incidence matrix and its properties, Symmetric BIBD, Resolvable BIBD, Affine Resolvable BIBD, Intra Block analysis, complimentary BIBD, Residual BIBD, Dual BIBD, Derived BIBD.

List of experiments (to be executed using Scientific Calculators and/or MS Excel)

1. Analysis of a CRD, RBD, LSD.
2. Analysis of an RBD with one missing observation.
3. Analysis of an LSD with one missing observation.
4. Analysis of 2^2 and 2^3 Factorial Experiments in CRD and RBD.
5. Analysis of a completely confounded two-level Factorial design in 2 blocks and 4 blocks.
6. Analysis of a partially confounded two-level Factorial design.
7. Intra Block analysis of a BIBD

Text Books:

T1. W G Cochran and G M Cox, Experimental Design, Asia Publishing House, 1959.
T2. M N Das and N C Giri, Design and Analysis of Experiments, Wiley Eastern Ltd., 1986.
T3. A M Gun, M K Gupta, and B Dasgupta, Fundamentals of Statistics, Vol. II, 9th Edition World Press, Kolkata, 2008.
T4. S C Gupta, and V K Kapoor, Fundamentals of Applied Statistics, 4th edition, Sultan Chand & Sons, 2008.

Reference Books:

R1. Kempthorne, The Design and Analysis of Experiments, John Wiley, 1965.
R2. D C Montgomery, Design and Analysis of Experiments, John Wiley, 2008.

Evaluation:

Mode of Evaluation	Hybrid			
Weightage	Comprehensive and Continuous Assessment		End Semester Examination	
	Theory	Practical	Theory	Practical
	25%	25%	25%	25%

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	3	2	2	1	1	1	1	1	1	2	2	3	3	2
CO2	3	3	2	2	2	2	2	2	1	1	3	2	3	3	2
CO3	3	3	3	2	3	3	2	2	2	1	2	3	3	3	3
CO4	3	3	3	3	3	3	2	2	2	2	2	3	3	3	3
CO5	3	2	3	3	3	2	2	2	2	2	2	3	3	3	3
CO6	3	2	3	3	3	3	2	3	2	3	3	3	3	3	3
	3.00	**2.67**	**2.67**	**2.50**	**2.50**	**2.33**	**1.83**	**2.00**	**1.67**	**1.67**	**2.33**	**2.67**	**3.00**	**3.00**	**2.67**

Course Title	**Research Methodology**
Course Code	**MTH450**
Credit	**4**
Contact Hours (L-T-P)	**3-1-0**
Course Type	**Theory**

Course Objectives

To equip students with a comprehensive understanding of research methodologies to formulate effective research questions and objectives. It emphasizes the improvement of scientific writing and communication skills for creating persuasive research papers, proposals, and presentations. Students will explore ethical considerations, including the avoidance of scientific misconduct and the promotion of responsible conduct, to maintain integrity in research. Additionally, the course will focus on effectively utilizing indexing and citation databases and mastering tools for scientific writing and reference management to ensure accurate documentation of sources.

Course Outcomes

On completion of this course, the students will be able to

CO1: Explain the nature, significance, and different approaches to research, differentiating between qualitative and quantitative methods.

CO2: Describe and apply ethical considerations and principles of responsible conduct in research.

CO3: Demonstrate the ability to structure research papers, proposals, and presentations effectively using appropriate scientific writing techniques.

CO4: Analyze scientific misconduct issues like falsification, fabrication, and plagiarism, and suggest preventive measures.

CO5: Integrate knowledge of databases, research metrics, and citation tools to evaluate the impact and quality of research outputs.

CO6: Assess the importance of intellectual property rights, open access, and research ethics in ensuring scientific integrity.

Course Description:

This course on Research Methodology covers essential concepts in research, including various approaches, ethical considerations, and the formulation of research questions. Students will learn to structure research papers, improve scientific writing, and deliver presentations. The course also addresses ethics in research, scientific misconduct, and explores indexing databases, research metrics, and tools like LaTeX for effective writing.

Course Content:

Module I: Introduction to Research Methodology:

Understanding the nature and significance of research, Different approaches to research - Basic, Applied, Interdisciplinary, Multidisciplinary, difference between quantitative and qualitative research, identifying research objectives and formulating research questions, exploring ethical considerations and responsible conduct of research.

Module II: Research communication and scientific writing:

Structuring and organizing research papers and reports, enhancing scientific writing skills for research articles and proposals, preparing effective research presentations (oral and poster), developing communication skills for scientific conferences and collaborations

Module III: Ethics in research:
Definition, moral philosophy, scientific conduct, intellectual honesty, and research integrity, scientific misconducts, falsification, fabrication and plagiarism, Redundant publications, duplication, overlapping publications, and salami slicing.

Module IV:
Databases, Indexing databases, Citation databases: Web of Science, Scopus, etc. Research Metrics, Impact factor, SNIP, SJR, IPP, Cite Score, h-index, g-index, i10 index, altmetrics. Scientific writing, References and citations (Medley, Jabref), plagiarism, intellectual property rights, copyrights, preprints (arXiv), open access.

Text Books

- **T1.** K. S. and Abbott, B. B., Research Design and Methods – A Process Approach, Bordens, Eighth Edition, McGraw-Hill, 2011.
- **T2.** C. R. Kothari, Research Methodology – Methods and Techniques, Second Edition, New Age International Publishers, 2004.
- **T3.** Bird A, Philosophy of Science, Routledge, 2006.
- **T4.** Praveen Chaddah, Ethics in Competitive Research: Do not get scooped; do not get plagiarized, 2018.
- **T5.** C. George Thomas, Research Methodology and Scientific Writing, Springer Nature, 2015.

Reference Books

- **R1.** Kumar, Ranjit, Research methodology: A step-by-step guide for beginners, 2018.
- **R2.** John W. Creswell and J. David Creswell. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches
- **R3.** Goldbort, Robert. Writing for science. Yale university press, 2006.
- **R4.** Davis, Martha, Kaaron Joann Davis, and Marion Dunagan. Scientific papers and presentations. Academic press, 2012.
- **R5.** Research Methods for Science, Michael P. Marder, Cambridge University Press, 2011

Evaluation:

Mode of Evaluation	Theory	
Weightage	Comprehensive and Continuous Assessment	End Semester Examination
	50%	50%

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	2	2	2	2	2	2	1	2	2	2	3	2	2	2
CO2	2	1	2	2	2	2	3	2	3	2	2	3	2	2	2
CO3	3	1	3	2	3	3	2	1	2	2	3	3	2	2	3
CO4	2	1	2	3	2	2	2	1	3	2	2	3	2	2	2
CO5	3	2	3	3	3	3	2	1	2	2	2	3	3	3	3
CO6	2	1	2	2	2	2	3	2	3	2	2	3	2	2	2
	2.50	**1.33**	**2.33**	**2.33**	**2.33**	**2.33**	**2.33**	**1.33**	**2.50**	**2.00**	**2.17**	**3.00**	**2.17**	**2.17**	**2.33**

Course Title	**Statistical Quality Control**
Course Code	**SDS451**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives:

- To introduce the concept and dimensions of quality, and trace the historical evolution of quality control and improvement.
- To familiarize students with key quality systems and standards, including ISO standards and quality registration processes.
- To equip students with knowledge of Statistical Process Control (SPC) tools and techniques for monitoring and improving process quality.
- To provide understanding of the construction and application of statistical control charts and process capability analysis.
- To introduce the principles of acceptance sampling plans and their practical applications in quality control.

Course Outcomes:

On completion of this course, the students will be able to:

CO1: Recall the definition, dimensions, and historical perspective of quality control, including the contributions of quality gurus and ISO standards.

CO2: Explain the principles of Statistical Process Control (SPC) and the construction and statistical basis of 3-σ control charts.

CO3: Apply statistical methods to construct and interpret control charts for variables and attributes, including pattern analysis and process capability estimation.

CO4: Compare between control charts for variables and attributes and analyze the effectiveness of statistical control tools.

CO5: Evaluate the effectiveness of single and double acceptance sampling plans using performance measures such as OC, AOQ, AOQL, ASN, and ATI.

CO6: Design and implement statistical quality control plans for real-world processes using tools like SPC charts and sampling plans, and interpret the results using software tools.

Course Description:

This course provides an in-depth understanding of quality management principles, focusing on the historical evolution of quality control from World War II to modern practices. It covers the contributions of Quality Gurus and explores key quality systems, including ISO standards. The course introduces Statistical Process Control (SPC), teaching students how to use control charts for monitoring process variability and improving quality. It also delves into acceptance sampling plans, offering insights into their principles and practical applications in quality inspection. By the end of the course, students will be equipped with essential tools and knowledge for effective quality management and process improvement.

Course Content:

Module I: Quality:

Definition, dimensions of quality, historical perspective of quality control and improvements starting from World War II, historical perspective of Quality Gurus and Quality Hall of Fame. Quality system and standards: Introduction to ISO quality standards, Quality registration. Statistical Process Control - Seven tools of SPC, chance and assignable Causes of quality variation. Statistical Control Charts- Construction and Statistical basis of 3-σ Control charts, Rational Sub-grouping

Module II: Control charts for variables

X-bar & R-chart, X-bar & s-chart. Control charts for attributes:
np-chart, p-chart, c-chart, and u-chart. Comparison between control charts for variables and control charts for attributes. Analysis of patterns on control chart, estimation of process capability.

Module III: Acceptance sampling plan

Principle of acceptance sampling plans. Single and Double sampling plan their OC, AQL, LTPD, AOQ, AOQL, ASN, ATI functions with graphical interpretation, use and interpretation of Dodge and Romig's sampling inspection plan tables.

Statistical Quality Control (PRACTICAL/LAB. WORK)

List of Practical

1. Construction and interpretation of statistical control charts
   - X-bar & R-chart
   - X-bar & s-chart
   - np-chart
   - p-chart
   - c-chart
   - u-chart

2. Single sample inspection plan: Construction and interpretation of OC, AQL, LTPD, ASN, ATI, AOQ, AOQL curves
3. Calculation of process capability and comparison of 3-sigma control limits with specification limits.

Text Books:

- **T1.** Goon A.M., Gupta M.K. and Dasgupta B. (2002): Fundamentals of Statistics, Vol. I & II, 8th Edn. The World Press, Kolkata.
- **T2.** Mukhopadhyay, P (2011): Applied Statistics, 2nd edition revised reprint, Books and Allied(P) Ltd.

Reference Books:

- **R1.** Montogomery, D. C. (2009): Introduction to Statistical Quality Control, 6th Edition, Wiley India Pvt. Ltd.

R2. Montogomery, D. C. and Runger, G.C. (2008): Applied Statistics and Probability for Engineers, 3rd Edition reprint, Wiley India Pvt. Ltd.

Evaluation:

Mode of Evaluation	Hybrid			
Weightage	Comprehensive and Continuous Assessment		End Semester Examination	
	Theory	Practical	Theory	Practical
	25%	25%	25%	25%

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	3	2	2	1	1	1	1	1	1	2	2	3	3	2
CO2	3	3	2	2	2	2	2	2	1	1	3	2	3	3	2
CO3	3	3	3	2	3	3	2	2	2	1	2	3	3	3	3
CO4	3	3	3	3	3	3	2	2	2	2	2	3	3	3	3
CO5	3	2	3	3	3	2	2	2	2	2	2	3	3	3	3
CO6	3	2	3	3	3	3	2	3	2	3	3	3	3	3	3
	3.00	**2.67**	**2.67**	**2.50**	**2.50**	**2.33**	**1.83**	**2.00**	**1.67**	**1.67**	**2.33**	**2.67**	**3.00**	**3.00**	**2.67**

Course Title	**Data Analytics for Finance**
Course Code	**SDS452**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives:

The objective of the courses are as follows:

- Develop a strong foundational understanding of financial analytics, its relevance in modern finance, and explore recent trends in data-driven financial decision-making
- Gain proficiency in reading, cleaning, and managing financial data, including handling missing values, identifying outliers, and performing exploratory data analysis with a focus on time series data
- Learn to apply key financial principles such as time value of money, risk and return, and asset valuation using Python programming, including implementing models like the Binomial and Black-Scholes for pricing options
- Understand the tools and techniques for detecting financial fraud, particularly credit card fraud, and gain experience with valuation analytics, focusing on real-world applications
- Learn how to evaluate portfolio return and risk, apply the Markowitz Optimization process, and assess portfolio performance using advanced financial metrics and models
- Acquire the skills to model stock price behavior using stochastic processes, and understand the concept of implied volatility for more informed decision-making in financial markets
- Apply the knowledge and techniques learned in the course to solve practical, real-world financial problems, equipping students with the skills necessary for a career in financial analysis or investment management.

Course Outcomes:

On completion of this course, the students will be able to:

CO1: Define and describe the concepts of financial analytics, including its relevance, scope, and recent trends.

CO2: Interpret and summarize financial data, including identifying missing data, outliers, and conducting exploratory analysis of time-series data.

CO3: Apply Python programming to perform basic financial computations such as time value of money, asset valuation, and risk estimation.

CO4: Analyze financial fraud scenarios and predict credit card fraud using appropriate data-driven methodologies.

CO5: Evaluate portfolio performance using Markowitz optimization, risk-return analysis, and performance metrics.

CO6: Develop integrated financial analytics solutions to address complex financial decision-making problems by synthesizing data and modern tools.

Course Description:

This course provides an in-depth exploration of financial analytics, equipping students with the tools and techniques necessary to analyze, interpret, and make data-driven decisions in the finance industry. Students will learn to manage and manipulate financial data, apply financial models using Python, and gain hands-on experience in addressing real-world financial challenges. The course begins with an introduction to financial analytics, its relevance, and recent trends in data-driven finance. Students will then delve into techniques for wrangling financial data, including handling missing values, outliers, and performing exploratory analyses, with a particular focus on time series data. In subsequent modules, students will gain a foundational understanding of finance concepts through Python, exploring time value of money, risk and return, asset valuation models such as the Binomial and Black-Scholes models, and the estimation of factor betas. The course also covers essential topics in financial fraud detection and prediction, with practical examples like credit card fraud analytics. The final module focuses on portfolio analytics, where students will learn to assess the return and risk of portfolios, apply the Markowitz optimization process, and evaluate portfolio performance using key financial metrics. By the end of the course, students will be able to analyze financial data, construct and evaluate investment portfolios, and apply advanced financial models to solve practical problems in the field of finance.

Course Content:

Module I

Introduction to Financial Analytics: Analytics in finance, Data driven finance, relevance and scope financial analytics, recent trends in financial analytics, data in Finance: types of financial data.

Module II

Wrangling Financial data: Reading and check financial data, managing missing data and outliers, exploratory analysis of financial data, analysis of time series data.

Module III

Understanding basic finance using Python: Time value of money, risk and return, estimating factor betas; Asset valuation: valuing a bond, share and option; Binomial option pricing model, valuing option using Black-Scholes model, stochastics process for stock prices, implied volatility.

Module IV

Financial fraud analytics, predicting credit card frauds, valuation analytics.

Module V

Portfolio analytics: Return and risk of a portfolio, Markowitz Optimization process, Portfolio performance evaluation.

Text Books

1. Mohanty, P., (2023): Financial Analytics, Wiley India pvt. Ltd.
2. Alexander, C. (2008). Quantitative methods in finance. Wiley.

Reference Book:

1. Alòs, E., Merino, R. (2023): Introduction to Financial Derivatives with Python, CRC Press.
2. Koop, G., (2006): ANALYSIS OF FINANCIAL DATA, John Wiley & Sons Ltd

3. Skoglund, J., & Chen, W. (2015). Financial risk management: Applications in market, credit, asset and liability management and firmwide risk. John Wiley & Sons.

Practical to be perform on computer using R/Python:

1. Financial data exploration (Individual)
2. Detecting patterns in financial statements (individual)
3. Predicting corporate bankruptcy (Group)
4. Develop valuation function and estimating implied returns (Individual)
5. Portfolio optimization and evaluation. (Group)
6. Predicting the stock returns using neural network model (Group)

Evaluation:

Mode of Evaluation	**Hybrid**			
Weightage	**Comprehensive and Continuous Assessment**		**End Semester Examination**	
	Theory	**Practical**	**Theory**	**Practical**
	25%	**25%**	**25%**	**25%**

Mapping COs with POs & PSOs:

CO/PO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**	**PSO1**	**PSO2**	**PSO3**
CO1	3	2	1	1	2	1	1	1	1	1	2	2	2	2	2
CO2	3	3	2	2	3	2	2	1	1	1	2	2	2	2	3
CO3	3	3	3	2	3	3	1	1	1	2	2	3	3	3	3
CO4	3	2	3	3	3	3	3	2	1	2	2	3	3	3	3
CO5	3	3	3	3	3	3	2	2	1	2	3	3	3	3	3
CO6	3	3	3	3	3	3	3	2	2	3	3	3	3	3	3
	3.00	**2.67**	**2.50**	**2.33**	**2.83**	**2.50**	**2.00**	**1.50**	**1.17**	**1.83**	**2.33**	**2.67**	**2.67**	**2.67**	**2.83**

Course Title	**Artificial Neural Network**
Course Code	**SDS405**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives:

- Develop the ability to train and apply single-layer and multi-layer neural networks using supervised learning techniques.
- Equip students with the skills to solve binary, multiclass classification, and regression problems using neural networks.
- Enable students to tackle data-driven challenges by addressing issues like vanishing gradients, overfitting, and computational load in deep neural networks.
- Foster the ability to design and implement full Convolutional Neural Network (CNN) architectures for image processing tasks.

Course Outcomes:

On completion of this course, the students will be able to:

CO1: Recall the structure and functions of biological and artificial neural networks, activation functions, and neural network architectures.

CO2: Explain supervised learning methods for single-layer and multi-layer neural networks, including training algorithms such as the delta rule and back-propagation.

CO3: Apply neural networks for binary and multiclass classification, regression, and learning from data using a two-layer neural network.

CO4: Analyze challenges in deep learning, such as vanishing gradients, overfitting, and computational load, and demonstrate the use of techniques like ReLU and dropout for improvement.

CO5: Evaluate the architecture and functionality of convolutional neural networks (CNNs), including convolution and pooling layers, strides, and full CNN architecture.

CO6: Design and implement solutions for real-life classification, regression, and pattern recognition problems using neural networks and CNNs with R/Python/MATLAB.

Course Description:

This course offers a comprehensive introduction to neural networks and deep learning, starting with the foundations of biological and artificial neural networks. Students will explore various types of activation functions and neural network architectures, with a focus on supervised learning and the training of single and multi-layer networks using techniques like Stochastic Gradient Descent (SGD) and the Back-Propagation Algorithm. The course covers neural networks for classification and regression, as well as few advanced topics in deep learning. It also focused in-depth study of Convolutional Neural Networks (CNNs), including their architecture, convolutional and pooling layers, and practical applications in computer vision.

Course Content:

Module I:

Neural network: introduction, understanding of a biological neural network, exploring the artificial neuron, types of activation function, architecture of neural network, supervised learning of a neural network, training of a single-layer neural network: delta rule, generalized delta rule, batch, and mini batch, stochastic gradient descent

Module II:

Training of multi-layer neural network: back-propagation algorithm, cost function and learning rule, cross entropy function

Module III:

Neural network for classification: binary classification, multiclass classification, neural networks for regression, two-layer neural network, learning the network from data.

Module IV:

Deep learning: improvement of the deep neural network, vanishing gradient, overfitting, computational load, relu and dropout.

Module V:

Convolutional neural network, architecture of convnet, convolution layer, pooling layer, condensing information with strides, full CNN architecture.

List of Experiments using R/Python/MATLAB:

1. Data fitting, clustering and pattern recognition using neural network
2. Solve a regression problem using ANN
3. Solve a classification problem using ANN
4. Solve various real life examples using CNN

Books Recommended:

1. B.Yegnanarayana, Artificial Neural Networks, Prentice Hall of India.
2. Satish Kumar, Neural Networks – A Classroom Approach, Tata McGraw-Hill.
3. S.Haykin, Neural Networks – A Comprehensive Foundation, Prentice Hall.
4. Phil Kim, MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence, Apress.
5. Kevin L. Priddy, Paul E. Keller – Artificial neural networks: An Introduction - SPIE Press, 2005
6. Mohammad H. Hassoun – Fundamentals of artificial neural networks - MIT Press ,1995

Evaluation:

Mode of Evaluation	Hybrid			
Weightage	Comprehensive and Continuous Assessment		End Semester Examination	
	Theory	Practical	Theory	Practical
	25%	25%	25%	25%

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	3	2	1	1	2	1	1	1	1	2	2	2	3	2
CO2	3	3	2	2	2	3	2	1	1	1	3	2	3	3	3
CO3	3	3	3	3	3	3	2	1	2	1	2	3	3	3	3
CO4	3	3	3	3	3	3	2	2	2	2	2	3	3	3	3
CO5	3	3	3	3	3	3	2	2	2	2	2	3	3	3	3
CO6	3	3	3	3	3	3	3	2	2	3	3	3	3	3	3
	3.00	**3.00**	**2.67**	**2.50**	**2.50**	**2.83**	**2.00**	**1.50**	**1.67**	**1.67**	**2.33**	**2.67**	**2.83**	**3.00**	**2.83**

Course Title	**Data Engineering**
Course Code	**SDS453**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives:

- Understand the Big Data Platform and its Use cases
- Provide an overview of Apache Hadoop
- Provide HDFS Concepts and Interfacing with HDFS
- Understand Map Reduce Jobs
- Provide hands on Hadoop Eco System
- Apply analytics on Structured, Unstructured Data.
- Exposure to Data Analytics with R.

Course Outcomes:

On completion of this course, the students will be able to:

CO1: Explain the fundamental concepts of Big Data, Hadoop ecosystem, and data analysis tools like R.

CO2: Illustrate the architecture, functionality, and data flow in Hadoop Distributed File System (HDFS).

CO3: Apply MapReduce concepts to solve simple data processing problems such as word frequency and temperature analysis.

CO4: Analyze the functionality and performance of Hadoop ecosystem tools like Pig, Hive, and HBase for various types of data operations.

CO5: Evaluate machine learning approaches for supervised and unsupervised learning in the context of Big Data Analytics using R.

CO6: Design and implement advanced MapReduce programs for complex data processing, such as matrix multiplication and weather analysis.

Course Description:

This course covers the fundamentals of big data and Hadoop, including HDFS, MapReduce, and the Hadoop ecosystem, while also introducing data analytics with R and machine learning techniques for effective data processing and analysis.

Course Content:

Module I: INTRODUCTION TO BIG DATA AND HADOOP

Types of Digital Data, Introduction to Big Data, Big Data Analytics, History of Hadoop, Apache Hadoop, Analyzing Data with Unix tools, Analyzing Data with Hadoop, Hadoop Streaming, Hadoop Echo System, IBM Big Data

Strategy, Introduction to Infosphere Big Insights and Big Sheets.

Module II: HDFS (Hadoop Distributed File System)
The Design of HDFS, HDFS Concepts, Command Line Interface, Hadoop file system interfaces, Data flow, Data
Ingest with Flume and Scoop and Hadoop archives, Hadoop I/O: Compression, Serialization, Avro and File-Based
Data structures.

Module III: Map Reduce
Anatomy of a Map Reduce Job Run, Failures, Job Scheduling, Shuffle and Sort, Task Execution, Map Reduce Types
and Formats, Map Reduce Features.

Module IV: Hadoop Eco System
Pig: Introduction to PIG, Execution Modes of Pig, Comparison of Pig with Databases, Grunt, Pig Latin, User Defined
Functions, Data Processing operators.
Hive: Hive Shell, Hive Services, Hive Metastore, Comparison with Traditional Databases, HiveQL, Tables, Querying
Data and User Defined Functions.
Hbase: HBasics, Concepts, Clients, Example, Hbase Versus RDBMS.
Big SQL: Introduction

Module V: Data Analytics with R
Machine Learning: Introduction, Supervised Learning, Unsupervised Learning, Collaborative Filtering. Big Data Analytics with Big R.

List of Experiments:
1. Install Apache Hadoop
2. Develop a MapReduce program to calculate the frequency of a given word in agiven file.
3. Develop a MapReduce program to find the maximum temperature in each year.
4. Develop a MapReduce program to find the grades of students.
5. Develop a MapReduce program to implement Matrix Multiplication.
6. Develop a MapReduce to find the maximum electrical consumption in each year given electrical consumption for each month in each year.
7. Develop a MapReduce to analyze weather data set and print whether the day is shinny or cool day.

Text Books

T1. Tom White “Hadoop: The Definitive Guide” Third Edit on, O’reily Media, 2012.
T2. Seema Acharya, Subhasini Chellappan, "Big Data Analytics" Wiley 2015.

Reference Books

R1. Michael Berthold, David J. Hand, "Intelligent Data Analysis”, Springer, 2007.
R2. Jay Liebowitz, “Big Data and Business Analytics” Auerbach Publications, CRC press (2013)

R3. Tom Plunkett, Mark Hornick, “Using R to Unlock the Value of Big Data: Big Data Analytics with Oracle R
R4. Enterprise and Oracle R Connector for Hadoop”, McGraw-Hill/Osborne Media (2013), Oracle press.
R5. Anand Rajaraman and Jef rey David Ulman, “Mining of Massive Datasets”, Cambridge University Press, 2012.
R6. Bill Franks, “Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics”, John Wiley & sons, 2012.
R7. Glen J. Myat, “Making Sense of Data”, John Wiley & Sons, 2007
R8. Pete Warden, “Big Data Glossary”, O’Reily, 2011.

Evaluation:

<table>
<tr><td>Mode of Evaluation</td><td colspan="4">Hybrid</td></tr>
<tr><td rowspan="3">Weightage</td><td colspan="2">Comprehensive and Continuous Assessment</td><td colspan="2">End Semester Examination</td></tr>
<tr><td>Theory</td><td>Practical</td><td>Theory</td><td>Practical</td></tr>
<tr><td>25%</td><td>25%</td><td>25%</td><td>25%</td></tr>
</table>

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	2	1	1	2	3	1	1	1	2	2	3	3	3	3
CO2	3	3	2	2	3	3	1	1	1	2	2	3	3	3	3
CO3	2	2	3	3	3	3	2	1	1	3	2	3	3	3	3
CO4	3	2	3	3	3	3	2	2	2	3	3	3	3	3	3
CO5	3	3	3	3	3	3	3	2	2	3	3	3	3	3	3
CO6	3	2	3	3	3	3	2	2	2	3	3	3	3	3	3
	2.83	**2.33**	**2.50**	**2.50**	**2.83**	**3.00**	**1.83**	**1.50**	**1.50**	**2.67**	**2.50**	**3.00**	**3.00**	**3.00**	**3.00**

Course Title	**Practical Financial Econometric**
Course Code	**SDS454**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives

This course aims to equip students with advanced analytical skills in finance and econometrics. Students will learn to analyze portfolio characteristics using single and multi-factor models, apply classical volatility and correlation models, evaluate time series models and their cointegration, and develop advanced econometric techniques such as quantile regression for parameter estimation and inference. Through practical applications, students will gain a comprehensive understanding of these essential concepts in financial analysis.

Course Outcomes

CO1: Recall the fundamental concepts of factor models, including single-factor and multi-factor models, and their applications in estimating portfolio risk and characteristics.

CO2: Interpret the classical models of volatility and correlation, including their significance in financial analytics and hedge fund volatility.

CO3: Apply time series models to analyze stochastic trends, random walks, and efficient market hypotheses using relevant mathematical and statistical tools.

CO4: Analyze the issues of multicollinearity in factor models and their implications for systematic risk estimation in real-world financial portfolios.

CO5: Evaluate econometric models, such as quantile regression, to assess and infer trends and variability in financial datasets.

CO6: Develop comprehensive financial analytics solutions by implementing advanced statistical models through practical mini-projects involving portfolio risk, volatility, and cointegration.

Course Descriptions

This course explores advanced topics in financial econometrics and portfolio management, covering four key areas. It begins with factor models, including single and multi-factor models for estimating portfolio characteristics and risk, and style attribution analysis. Next, it examines classical models of volatility and correlation, focusing on variance, covariance matrices, and forecasting techniques for hedge funds. The course then delves into time series models and cointegration, emphasizing estimation, prediction, and reconciliation with continuous time models. Finally, it covers advanced econometric techniques such as quantile regression and parameter estimation, equipping students with the skills to analyze and interpret complex financial data. By the end of this course, students will have a strong understanding of advanced econometric methods and their applications in portfolio management.

Course Content:

Module I:

Factor Models: Introduction, Single Factor Models, estimating Portfolio characteristics using OLS, Estimating Portfolio Risk using EWMA, Relationship between Beta, Correlation and relative

Volatility, Multi-factor Models of asset or portfolio returns, Style Attribution Analysis, General Formulation of Multi-factor Model.

Module II:
Classical models of volatility and correlation: Introduction, variance and volatility, volatility for hedge funds, covariance and correlation Matrices, equally weighted averages, Unconditional Variance and Volatility, Forecasting with equally weighted averages

Module III

Time Series Models and Cointegration: Review of Time Series Models, Inversion and the Lag Operator, Response to Shocks, Estimation and Prediction, Stochastic Trends, Random Walks and Efficient Markets, Integrated Processes, Reconciliation of Time Series and Continuous Time Models.

Module IV
Advanced Econometric Models: Quantile Regression, Parameter Estimation in Quantile Regression, Inference on Linear Quantile Regressions.

Practical to be conducted through various mini project work:
Project1: Case Study: i) Estimation of Fundamental Factor Models, ii) Estimating Systematic Risk for a Portfolio of US Stocks, iii) Multicollinearity: A Problem with Fundamental Factor Models.
Project2: Case Study: Volatility and Correlation of US Treasuries
Project3: Case Study: Cointegration Index Tracking in the Dow Jones Index
Project4: Case Study: Quantile Regression on FTSE 100 Index

Text Books

- **T1.** Alexander, C., Market Risk Analysis Volume II: Practical Financial Econometrics, John Wiley & Sons Ltd, 2008.
- **T2.** Chris Brooks, Introductory Econometrics for Finance, 2nd edn. Cambridge University Press, 2008.

Reference Books:

- **R1.** John Hull. Options, Futures, and Other Derivatives, Fifth Edition. Prentice Hall.
- **R2.** Alexander, C. (2008). Quantitative methods in finance. Wiley, 2009.
- **R3.** Skoglund, J., & Chen, W., Financial risk management: Applications in market, credit, asset and liability management and firmwide risk. John Wiley & Sons, 2015.

Evaluation:

<table>
<tr><td>Mode of Evaluation</td><td colspan="4">Hybrid</td></tr>
<tr><td rowspan="3">Weightage</td><td colspan="2">Comprehensive and Continuous Assessment</td><td colspan="2">End Semester Examination</td></tr>
<tr><td>Theory</td><td>Practical</td><td>Theory</td><td>Practical</td></tr>
<tr><td>25%</td><td>25%</td><td>25%</td><td>25%</td></tr>
</table>

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	2	1	1	2	1	0	0	1	1	1	2	2	3	2
CO2	3	2	2	1	3	2	0	0	1	1	2	2	3	2	3
CO3	3	3	3	2	3	3	1	1	1	2	2	3	3	3	3
CO4	3	3	3	3	3	3	1	1	1	2	2	3	3	3	3
CO5	3	3	3	3	3	3	2	1	1	2	3	3	3	3	3
CO6	3	3	3	3	3	3	2	2	2	3	3	3	3	3	3
	3.00	**2.67**	**2.50**	**2.17**	**2.83**	**2.50**	**1.00**	**0.83**	**1.17**	**1.83**	**2.17**	**2.67**	**2.83**	**2.83**	**2.83**

Course Title	**Data Management and Data Warehousing**
Course Code	**SDS455**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives:

- To introduce the foundational knowledge of data warehousing components and architecture, emphasizing tools for data extraction, transformation, and OLAP for effective reporting.
- To be able to implement robust database systems through feasibility studies, normalization techniques, and the use of data dictionaries to ensure data integrity.
- To explore SQL query construction, including complex joins and subqueries, alongside the design of forms and reports for optimal data presentation and retrieval.

Course Outcomes:

On completion of this course, the students will be able to:

CO1: Identify the components and architecture of data warehousing and their applications in business analysis.

CO2: Explain database management concepts, including class diagrams, normalization, and data dictionary, and how they support database development.

CO3: Apply SQL queries to perform computations, create forms, generate reports, and handle procedural programming tasks effectively.

CO4: Analyze view management and data security principles in centralized and distributed database management systems.

CO5: Evaluate data mining techniques such as classification, clustering, and regression for analyzing complex datasets, including multimedia and web data.

CO6: Design and implement data preprocessing tasks, build data warehouses, and perform practical projects in data mining and warehousing using MATLAB or C programming.

Course Description:

This course provides a comprehensive exploration of data warehousing, database management systems, and advanced data mining techniques. Students will learn the architecture and components of data warehouses, database design principles, and the intricacies of SQL for querying and data manipulation. Emphasis will be placed on creating user-friendly interfaces, effective reporting, and ensuring data integrity and security. Additionally, the course covers the mining of various data types, including spatial and multimedia data, equipping students with the skills to derive actionable insights. Through practical projects, learners will gain hands-on experience relevant to real-world data analysis scenarios.

Course Content:

Module I

Data Warehousing and Business Analysis: - Data warehousing Components —Data Warehouse Architecture – DBMS Schemas for Decision Support – Data Extraction, Cleanup, and

Transformation Tools –Metadata – reporting – Query tools and Applications – Online Analytical Processing (OLAP) – OLAP and Multidimensional Data Analysis.

Module II

Database Management Systems – Feasibility Study – Class Diagrams – Data Types – Events – Normal Forms – Integrity – Converting Class Diagrams to Normalized Tables – Data Dictionary.

Module III

Query Basics – Computation Using Queries – Subtotals and GROUP BY Command – Queries with Multiple Tables – Subqueries – Joins – DDL & DML – Testing Queries, Effective Design of Forms and Reports – Form Layout – Creating Forms – Graphical Objects – Reports – Procedural Languages – Data on Forms – Programs to Retrieve and Save Data – Error Handling

Module IV

View Management, Views in Centralized DBMSs, Views in Distributed DBMSs, Maintenance of Materialized Views, Data Security Discretionary Access Control, Multilevel Access Control, Distributed Access Control, Semantic Integrity Control, Centralized Semantic Integrity Control, Distributed Semantic Integrity Control.

Module V

Mining Object, Spatial, Multimedia, Text and Web Data: Multidimensional Analysis and Descriptive Mining of Complex Data Objects – Spatial Data Mining – Multimedia Data Mining – Text Mining – Mining the World Wide Web.

Practical/Lab to be performed on a computer using MATLAB/C programming:

1. Build Data Warehouse
2. Perform data preprocessing tasks and Demonstrate performing association rule mining on data sets
3. Demonstrate performing classification on data sets
4. Demonstrate performing clustering on data sets
5. Demonstrate performing Regression on data sets
6. Task 1: Credit Risk Assessment. Sample Programs using German Credit Data
7. Task 2: Sample Programs using Hospital Management System
8. Simple Project on Data Pre-processing Data Warehousing and Mining Lab

Text Books:

T1. Alex Berson and Stephen J. Smith "Data Warehousing, Data Mining & OLAP", Tata McGraw – Hill Edition, Tenth Reprint 2007.

T2. M. Tamer Ozsu, Principles of Disrtibuted Database Systems, Springer 3rd Edition.

T3. Saeed K. Rahimi, Frank S. Haug, Distributed Database Management system.

Reference Books:

R1. K.P. Soman, Shyam Diwakar and V. Ajay "Insight into Data mining Theory and Practice", Easter Economy Edition, Prentice Hall of India, 2006.

R2. G. K. Gupta "Introduction to Data Mining with Case Studies", Easter Economy Edition, Prentice Hall of India, 2006.

R3. Pang-Ning Tan, Michael Steinbach and Vipin Kumar “Introduction to Data Mining”, Pearson Education, 2007.

Evaluation:

Mode of Evaluation	Hybrid			
Weightage	Comprehensive and Continuous Assessment		End Semester Examination	
	Theory	Practical	Theory	Practical
	25%	25%	25%	25%

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	2	1	1	2	1	0	0	0	1	1	2	2	2	2
CO2	3	2	2	1	3	2	0	0	0	1	2	2	3	2	3
CO3	3	2	3	2	3	3	1	0	1	2	2	3	3	3	3
CO4	3	3	3	3	3	3	1	1	1	2	2	3	3	3	3
CO5	3	3	3	3	3	3	2	1	1	2	3	3	3	3	3
CO6	3	3	3	3	3	3	2	2	2	3	3	3	3	3	3
	3.00	**2.50**	**2.50**	**2.17**	**2.83**	**2.50**	**1.00**	**0.67**	**0.83**	**1.83**	**2.17**	**2.67**	**2.83**	**2.67**	**2.83**

Course Title	Project/Dissertation
Course Code	**SDS499**
Credit	**12**
Contact Hours (L-T-P)	**- - -**
Course Type	**Dissertation**

Course Objectives:

- To address the real-world problems and find the required solution.
- To fabricate and implement the mini project intended solution for project-based learning
- To improve the team building, communication and management skills of the students

Course Outcomes

On completion of this course, the students will be able to

CO1: Describe the fundamental principles of statistics and data science applied in project work.

CO2: Explain the importance of statistical analysis in diverse fields such as data science, agriculture, finance, and industry.

CO3: Apply statistical theories and tools to real-world datasets to derive meaningful conclusions.

CO4: Analyze statistical problems using modern tools and interpret results for reliable decision-making.

CO5: Evaluate statistical methods and their applications in designing solutions to societal and industrial challenges.

CO6: Develop a comprehensive statistical project that integrates theoretical knowledge and practical skills for career advancement.

Course Description:

The role of project dissertation in Statistics and Data Science is very important. Project helps a student to explore and strengthen the understanding of fundamentals through practical application of theoretical concepts. Statistics is the basis of many areas including data science, agriculture, finance, industry. Everything around us needs statistical analysis for acceptance. Project allows a student to apply Statistical theory in practical purpose to draw some valid conclusions. It acts like a beginner's guide to do larger projects later in their career. It not just affects the grades of the program but also matter a lot for good CV/Resume. So before choosing the minor and major project, you should explore the options and pick the correct domain where the opportunities are immense.

Evaluation:

Mode of Evaluation	Presentation and Report
Weightage	End Semester Examination
	100%

Mapping COs with POs & PSOs:

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	2	1	-	1	-	-	-	-	-	1	2	2	1	1
CO2	3	3	2	-	-	-	2	1	-	-	-	2	3	2	2
CO3	3	3	3	2	3	2	2	2	-	2	2	2	3	3	3
CO4	3	3	3	3	3	3	2	2	-	2	2	3	3	3	3
CO5	3	3	3	3	3	3	3	3	1	3	2	3	3	3	3
CO6	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	3.00	**2.83**	**2.50**	**2.75**	**2.60**	**2.75**	**2.40**	**2.20**	**2.00**	**2.50**	**2.00**	**2.50**	**2.83**	**2.50**	**2.50**