Adamas University
School of Basic and Applied Sciences
Department of Mathematics

Syllabus of

B.Sc (Statistics and Data Analytics)

Programme Code: MTH3402
Duration: 4 Years Full Time
Academic Year: 2023-24

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 (Statistics and Data Analytics):

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 (Statistics and Data Analytics):

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.

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. Define the terms related to set theory, relation, mapping, and function

CO2. Illustrate the concept of probability, Bayes' theorem, and its applications

CO3. Solve the problems related to probability distribution, expectation and moments

CO4. Build the fundamental knowledge of one- and two-dimensional random variables and related properties.

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, Statistical Probability and Axiomatic definition of probability, Applications.

Laws of addition and multiplication, Conditional Probability, Theorem of Total Probability, Bayes' theorem and its applications, Independent Events.

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, dispersion, skewness, kurtosis, 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, conditional expectation and variance.

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%

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. Demonstrate tabular and diagrammatic representation of frequency distribution of discrete and continuous data.

CO2. Illustrate descriptive measures for univariate data.

CO3. Illustrate descriptive measures for bivariate data.

CO4. Explain consistency and associations between attributes.

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, statistical Population and Sample. Different types of data, Nominal, Ordinal, Interval and Ratio scales of measurement, tabular and graphical frequency

distributions of data, cumulative frequency distributions, Stem, Leaf displays and their graphical representations.

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, simple Correlation, simple linear Regression, principle of least squares, 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, 8th Edn. The World Press, Kolkata.

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

T3. Yule G.U. and Kendall M.G. 1994. An Introduction to the theory of Statistics, 14th Edn. Universal Book Stall, Delhi.

T4. Das, N. G. 2017. Statistical Methods (Combined edition volume 1 & 2), McGraw Hill Education; 1st edition.

T5. S. K. De and S. Sen A. Banerjee.1999. MATHEMATICAL PROBABILITY, U. N. Dhur & Sons Pvt. Ltd.; 4th edition

Reference Books:

R1. Nguyen, H. T., & Rogers, G. S. 1989. Fundamentals of Mathematical Statistics: Probability for Statistics. Springer Science & Business Media.

R2. Agresti, A. 2010. Analysis of Ordinal Categorical Data, 2nd Edition, Wiley.

Evaluation:

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

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. Construct **t**he different generating functions.
CO2. Apply the concepts of different probability inequalities, different modes of convergence and central limit theorem to solve various problems.
CO3. Analyze the different discrete and continuous probability distributions, along with their properties and applications.
CO4. Solve the real-life applications using correlation and regression.

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, Cumulant Generating Function and Characteristic Function, Uniqueness, and Inversion theorems (without proof) 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. Limiting cases

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

Module IV:
Transformation of Random variables, Sum and product of Random Variables, Principal of Least squares, Correlation, 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	
	Theory	Practical	Theory	Practical
	25%	25%	25%	25%

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. Demonstrate the basic concepts of Sampling Theory and Sampling Distributions of a Statistic.

CO2. Develop knowledge on the exact Sampling Distributions - χ^2, t and F.

CO3. Illustrate the method of finding Sampling Distributions from Bivariate Normal population.

CO4. Solve problems related to Order Statistics and distributions of Sample Median and Sample Range.

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 & parameter, random sample & statistic, sampling distribution of a statistic and its applications, sampling fluctuations & standard error of a statistic, sampling distributions for functions of random variables – cdf technique, generating functions technique, transformation of variables technique with examples

Module II:
Exact Sampling Distributions: χ^2 distribution and its p.d.f. with 'n' degrees of freedom (d.f.), mean, variance, M.G.F., properties of χ^2 distribution. sampling distribution of sample mean and variance for a normal population, Student's and Fisher's t-distributions: p.d.f., mean, variance and limiting

form of 't' distribution. Snedecor's F-distribution: p.d.f., mean, variance, distribution of 1/F (n1, n2), Relationship between t, F and χ^2 distributions, applications of t, F and $\chi 2$ distribution in testing of significance problems.

Module III:
Bivariate normal population and order statistics: distributions of sample means, sample variances and sample correlation coefficient (null case) of a random sample from a bivariate normal population, distribution of the simple regression coefficient

Module IV:
Order statistics: introduction, distribution of the r^{th} order statistic, smallest and largest order statistics, joint distribution of order statistics, distribution of sample median and sample range, applications

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, 4^{th} Edn., The World Press, Kolkata.

T2. Rohatgi V. K. and Saleh A.K. Md. E. 2009. An Introduction to Probability and Statistics, 2^{nd} 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, 3^{rd} 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	
	Theory	Practical	Theory	Practical
	25%	25%	25%	25%

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 solid foundation in estimation theory and hypothesis testing, including key concepts and methodologies.
- Apply statistical techniques to real-world data, formulating and evaluating hypotheses effectively.
- Gain proficiency in advanced statistical methods, including likelihood ratio tests and the Rao-Blackwell and Lehmann-Scheffé theorems.
- Enhance critical thinking and problem-solving abilities, preparing students for further studies and practical applications in statistics.

Course Outcomes:

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

CO1: Interpret the concepts of Estimation, Unbiasedness, Sufficiency, Consistency and Efficiency, Complete Statistic, Minimum Variance Unbiased Estimator and Minimum Variance Bound.

CO2: Construct the Statistical Hypothesis and error analysis with calculation of Significance Level & Power of test

CO3: Develop the concepts of MP and UMP Tests and applications of Neyman-Pearson Lemma.

CO4: Make use of different statistical tools and techniques for statistical inference.

Course Description:

The course aims to provide students with a comprehensive understanding of estimation theory and hypothesis testing, equipping them with essential statistical tools. Students will learn to formulate and evaluate hypotheses while applying various statistical tests to real-world data. The curriculum emphasizes the comparison of estimators and the application of advanced techniques such as likelihood ratio tests. Ultimately, the course prepares students for further studies in statistics and enhances their analytical skills for practical applications.

Course Content:

Module I: Estimation

Estimation theory: Introduction, classifications, point, confidence level, 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, methods of point estimation, applications, and interval estimation

.

Module II: Hypothesis

Elements of hypothesis testing: Null and alternative hypotheses, simple & composite hypotheses, critical region, type I and type II errors, level of significance, size, power, p-value. exact tests and confidence intervals: classical and p-value approaches, tests of significance related to discrete and continuous distributions, combination of probabilities in tests of significance.

Module III: Hypothesis testing

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

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

Sl. No.	Name of the Experiments
1.	Unbiased Estimators (including Unbiased but absurd Estimators).
2.	Maximum Likelihood Estimation.
3.	Estimation by the Method of Moments.
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.
10.	Most Powerful Critical Region (NP Lemma).
11.	Uniformly Most Powerful and Unbiased Critical Region.
12.	Power Curves.

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	
	Theory	Practical	Theory	Practical
	25%	25%	25%	25%

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. Solve system of linear equations and diagonalize the matrix

CO2. Explain the concept of basis, dimension, linear mapping, and inner product.

CO3. Explain the concept of multivariate data and multivariate distribution for inference

CO4. Apply the dimension reduction techniques and factor analysis for multivariate data

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.

Evaluation:

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

Course Title	**Exploratory Data Analysis and Visualization**
Course Code	**SDS203**
Credit	**4**
Contact Hours (L-T-P)	**3-0-2**
Course Type	**Hybrid**

Course Objectives:

To learn and apply the various tools and techniques for solving problems related to exploratory data analysis and visualization using programming languages such as R/Python/Excel/MATLAB, based on real-life problems.

Course Outcomes

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

CO1. Define concept of exploratory data analysis and data visualization, summary statistics, univariate & multivariate visualization and data abstraction.

CO2. Illustrate different types of variables, techniques of interpretation of the data, data import, data cleaning, exploratory graphics of a financial dataset and functional data and common visualization idioms.

CO3. Solve problems related to exploratory data analysis and data visualization.

CO4. Interpret the data and summary statistics.

CO5. Explain data summarization and the use of colour and Size in visualization.

CO6. Solve problems related to data abstraction, encoding Data, decision trees, k-nearest neighbours.

Course Description:

This course includes exploratory data analysis and data visualization, summary statistics, interpretation of the data, data import, data cleaning, exploratory graphics, data abstraction, encoding Data, decision trees, k-nearest neighbours. All the lectures will be devoted on discussions of basic theories and advanced topics, focusing on practical implementation of knowledge. Classes will be conducted by lecture as well as power point presentation, audio visual virtual lab session as per requirement. The tutorials will familiarize the students with practical problem-solving techniques led by the course coordinator. Students will strongly grab the basic concepts of the subject via exercise and discussions with the coordinator.

Course Content:

Module I:

Exploratory data Analysis fundamental: Introduction to Data Science, Data Science hype, getting past the hype, Datafication, Current landscape of perspectives, Introduction to exploratory data analysis (EDA) and data visualization, Philosophy of EDA, Basic steps in EDA, making sense of data, studying individuals, studying variables, software tools for EDA.

Module II:
Data Visualization: Data visualization introduction, Univariate visualization: quantitative and categorical distributions, common visualization idioms, line charts, bar chart, Scatter plot, pie chart, area chart, visualization of spatial data, Histogram, Lollipop chart.

Module III:
Data Transformation: Replace values, handling missing values, Discretization and binning, outlier detection and filtering, string manipulation, Scaling the data, normalization, Summary statistics (mean, median, mode, variance, standard deviation, Skewness, kurtosis, Quartiles), interpreting the data, data import, data cleaning, correspondence analysis, Data aggregation, Pivot table.

Module IV:
Dimensionality, Goodness of fit testing, Graphical methods, Variance-based data projections, Distance-based data projections, other projection methods, other methods, Data dredging, Whole dataset: peeking into high-dimensional spaces, clustering (basics only), Exploratory Graphics of a Financial Dataset, Visualizing Functional Data.

List of Experiments:
(Use R or Python/Excel/Matlab platform to perform the following experiments)

1. Creating data frames
2. Extracting data from data frames
3. Import and export data files
4. Data cleaning, Treatment of missing values
5. Descriptive statistics and graphics
6. Visualization using tools (plots, graphs and summary statistics)
7. Visualizing Functional Data.
8. Applications based on data visualization techniques
9. Understanding univariate analysis
10. Understanding bivariate analysis
11. Understanding multivariate analysis

- **Small Project based on EDA using real Data Analysis**

Text Books:

T1. F Husson, S Lê, and J Pagès. 2017. Exploratory multivariate analysis by example using R. Chapman and Hall/CRC.
T2. J W Tukey. 1977. Exploratory data analysis (Vol. 2).
T3. C H Chen, W K Härdle, & A Unwin (Eds.), 2007. Handbook of data visualization. Springer Science & Business Media,.
T4. Suresh Kumar, Mukhiya Usman Ahmed, Hands-On Exploratory Data Analysis with Python. Packt Publishing
T5. Van der M Loo, & E de Jonge, 2018. Statistical data cleaning with applications in R. John Wiley &Sons.

Reference Books:

R1. T W Anderson. 2003. An Introduction to Multivariate Statistical Analysis, 3rd Edn. John Wiley.

R2. A H Fielding, 2006. Cluster and classification techniques for the biosciences. Cambridge University Press.

R3. B Everitt, & T Hothorn, 2011. An introduction to applied multivariate analysis with R. Springer Science & Business Media.

R4. Yau, N. 2011. Visualize this: the Flowing Data guide to design, visualization, and statistics. John Wiley & Sons.

R5. Haider, M. Getting Started with Data Science: Making Sense of Data with Analytics. IBM Press.

Evaluation:

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

Course Title	**Introduction To Linear Programming and Game Theory**
Course Code	**MTH204**
Credit	**4**
Contact Hours (L-T-P)	**3-1-0**
Course Type	**Theory**

Course Objectives:

- 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: Find the mathematical form of real-world problems in the form of LPP and find the solution by using Graphical method

CO2: Illustrate the concept of Simplex method and Duality of LPP

CO3: Choose appropriate techniques for solving problems related to transportation, assignment and travelling salesman

CO4: Build up the concept of application procedures of LPP to Games and strategies

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: Introduction to LPP

Introduction, definition of linear programming problem (LPP), formation of LPP, graphical Method, basic solutions and basic feasible solution (BFS), degenerate and non-degenerate BFS, Hyperplane, convex set, extreme points, convex hull and convex polyhedron, supporting and separating hyperplane, reduction of a feasible solution (FS) to a BFS, optimality condition, unboundedness, alternate optima, infeasibility

Module II: Simplex method
Theory of simplex method, optimality and un-boundedness, the simplex algorithm, 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
Transportation problem and its mathematical formulation, north-west corner method, least cost method and Vogel approximation method for determination of initial basic feasible solution, assignment problem and its mathematical formulation, Hungarian method for solving assignment problem, travelling salesman problem

Module IV: Game Theory
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

Text Books:

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

Reference Books:

- **R1.** F.S. Hillier and G.J. Lieberman, 2009. Introduction to Operations Research, 9th Ed., Tata McGraw Hill, Singapore.
- **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, 2008. Linear Programming and Game Theory, U. N. Dhur and sons Pvt Ltd.

Evaluation:

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

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**: Explain the fundamental concepts of data analytics
- **CO2**: Choose appropriate data analysis techniques for handling large data
- **CO3**: Develop various machine learning algorithms used in data analytics process
- **CO4**: Demonstrate the visualization and present the inference using various tools
- **CO5**: Discuss the ethics surrounding privacy and algorithmic decision-making.

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%

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. Demonstrate the concepts of Gauss-Markov set-up

CO2. Make use of method of least squares in Simple and Multiple Regression Models

CO3. Construct Analysis of Variance Model to test the significance of several means.

CO4. Build model prediction using various statistical tools

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, Method of Least Squares, Gauss-Markov theorem, Estimation of error variance. Fundamental Theorems on Least Squares (statements only), 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. Estimation and Hypothesis Testing in case of Simple and Multiple Regression Models. 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 and Covariance

Definitions of Fixed, Random and Mixed Effect Models, Analysis of Variance and Covariance in One-way classified data for Fixed Effect Models, Analysis of Variance and Covariance (with one concomitant variable) 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. 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 Covariance of a One-way classified data with one Concomitant variable.
9	Analysis of Covariance of a Two-way classified data with one Concomitant variable.
10	Analysis of Variance of a One-way classified data for Random Effects Model.
11	Logistic and Poisson 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	
	Theory	Practical	Theory	Practical
	25%	25%	25%	25%

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. Acquire the knowledge of big data, data science components and tools.

CO2. Develop the fundamental concept of different machine learning algorithms, clustering and classification techniques and its real-life applications.

CO3. Built the basic knowledge of recommendation system, singular value decomposition, principal component analysis and clustering of graphs with applications.

CO4. Develop the concept of data visualization, privacy, security, ethics and role of next generation data scientist.

Course Description:

This course is designed to provide students a fundamental concepts 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%

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. Build the concepts of various Index Numbers and their uses.

CO2. Develop the concepts of different components and models of a Time Series data

CO3. Estimate the Trend and Seasonal components using different methods.

CO4. Utilize Moving Average and Auto Regressive Process in Forecasting

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%

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.** Understand the professional requirements for access to and success in the field.
- **CO2.** Apply techniques using different methods of applying skills and knowledge acquired in the classroom.
- **CO3.** Discover the work ethic and skills required for success in the field.

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

Course Title	**Survey Sampling**
Course Code	**SDS304**
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 fundamental differences between population and sample, and the principles of probability and non-probability sampling.
- Apply various sampling techniques, including simple random, stratified, and systematic sampling, to estimate population parameters and their variances.
- Analyse advanced estimation methods like ratio, regression, and cluster sampling, comparing their efficiency with standard sampling techniques.

Course Outcomes:

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

CO1. Explain key concepts of sampling, including population, sample, and various sampling methods such as SRS, stratified sampling, and cluster sampling.

CO2. Apply sampling techniques like SRS and stratified sampling to estimate population parameters.

CO3. **Analyze** the efficiency of different sampling methods (SRS, stratified, and systematic sampling) in various scenarios, particularly in the presence of linear trends.

CO4. **Evaluate** the precision and accuracy of ratio and regression methods of estimation compared to SRS, and determine the relative efficiency of cluster sampling.

Course Description:

This course covers key sampling techniques, including simple random, stratified, and systematic sampling, along with advanced methods like ratio, regression, and cluster sampling. Students will learn to estimate population parameters, their variances, and determine sample size. The course emphasizes the distinction between population and sample, as well as probability and non-probability sampling methods.

Course Content:

Module I

Concept of population and sample, complete enumeration versus sampling, sampling and non-sampling errors. 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 proportion, variances of these estimates, estimates of their variances and sample size determination.

Module II

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. Systematic Sampling: Technique, estimates of population mean and total, variances of these estimates (N=nxk). Comparison of systematic sampling with SRS and stratified sampling in the presence of linear trend and corrections.

Module III

Introduction to Ratio and regression methods of estimation, first approximation to the population mean and total (for SRS of large size), variances of these estimates and estimates of these variances, variances in terms of correlation coefficient for regression method of estimation and their comparison with SRS. Cluster sampling (equal clusters only) estimation of population mean and its variance, comparison (with and without randomly formed clusters). Relative efficiency of cluster sampling with SRS in terms of intra class correlation. Concept of sub sampling.

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. Cochran W.G. (1984): Sampling Techniques (3rd Ed.), Wiley Eastern. John B Fraleigh, First Course in Abstract Algebra, Pearson.

T2. Murthy M.N. (1977): Sampling Theory & Statistical Methods, Statistical Pub. Society, Calcutta**.**

T3. Des Raj and Chandhok P. (1998): Sample Survey Theory, Narosa Publishing House.

T4. Goon A.M., Gupta M.K. and Dasgupta B. (2001): Fundamentals of Statistics (Vol.2), World Press. Wiley & Sons, 2005.

Reference Books:

R1. Sukhatme,P.V., Sukhatme,B.V. Sukhatme,S. Asok,C.(1984). Sampling Theories of Survey with Application, IOWA State University Press and Indian Society of Agricultural Statistics S. K. Mapa, Higher Algebra, Sarat book house.

R2. Parimal Mukhopadhyay (2008): Theory and Methods of Survey Sampling, Prentice Hall India Learning Private Limited; 2nd edition

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%

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: Classify the various types of machine learning techniques and know their applications

CO2: Solve various classification & regression problem with real life examples

CO3: Compare classification and clustering techniques with performance metric

CO4: Develop the machine learning techniques to find the solution of various data driven problems.

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%

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. Describe the characteristics of database and the architecture of Database system.

CO2. Model the elements used in Entity- Relationship diagram.

CO3. Summarize relational model concept and illustrate the relational constraints.

CO4. Build Structured Query Language (SQL) and apply to query a database.

CO5. Define normalization for relational databases.

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%

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: Apply the concept of basic calculus and portfolio theory for simple financial problems.

CO2: Utilize the linear regression models and ANOVA in financial modeling.

CO3: Apply various numerical techniques to calculate the risk and bond yield in financial applications.

CO4: Explain the various types of risk, portfolios and related theories, capital asset pricing model.

CO5: Find eigen vectors and eigen values in applications to linear portfolios.

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%

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. Recall the requirements for the real-world problems.

CO2. Demonstrate the project successfully by visualizing, analysing, testing and interpreting.

CO3. Construct data analytical skills.

CO4. Examine the findings of the study conducted in the preferred domain

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

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: Interpret probability generating function and its use in stochastic process.
CO2: Explain different concepts related to discrete-time Markov chain.
CO3: Apply the knowledge of continuous-time Markov chain in real-world problem.
CO4: Apply the knowledge of queueing theory in real-wstaorld problem.

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:

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

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:

CO1. Explain the algorithms and functioning of advanced techniques and concepts such as deep learning, distance metric learning, and domain adaptation.

CO2. Analyze the advantages and limitations of the various algorithms

CO3. Evaluate their potential applications in different contexts.

CO4. Apply several public domain machine learning toolboxes to solve real-world problems.

CO5. Design and conduct experiments to evaluate machine learning models.

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%

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. Explain the characteristics of options, bonds and interest rates, the theoretical relationships between spot, forward, and futures prices in financial markets.

CO2. Analyze various option pricing models, such as the Binomial and Black-Scholes-Merton, and understand the concepts of hedging, volatility smiles and skews.

CO3. Explain the principles of asset pricing theories, the Capital Asset Pricing Model (CAPM), significance of beta for individual assets and portfolios.

CO4. Evaluate portfolio performance using different measures.

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%

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. Explain the underlying concepts of different experimental design

CO2. Utilize the Design of experiments for various problems

CO3. Analyze 2^n factorial experiments, using confounding techniques and interpreting factor interactions.

CO4. Examine the BIBD and their variants, understanding parameter relationships and conducting intra-block analysis.

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%

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.** Identify the nature and significance of several types of research, objectives for doing research.
- **CO2.** Develop skills in scientific writing and present research effectively.
- **CO3.** Apply ethical principles in research, ensuring responsible conduct and treatment of human subjects.
- **CO4.** Select the appropriate research quality journal based several criteria for publishing the research.

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%

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. Apply the seven SPC tools, distinguish between causes of variation, and construct 3-σ control charts using rational subgrouping.

CO2. Analyze and apply X-bar & R-charts, X-bar & s-charts, attribute control charts (np, p, c, u).

CO3. Compare control charts for variables and attributes, identify patterns, and estimate process capability.

CO4. Analyze and implement acceptance sampling plans by using Dodge and Romig's sampling inspection tables, interpret their key functions (OC, AQL, LTPD, etc.).

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%

Course Title	**Data Analytics Principles and Techniques**
Course Code	**SDS452**
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: Explain the fundamental concepts of data analytics
CO2: Choose appropriate data analysis techniques for handling large data
CO3: Develop various machine learning algorithms used in data analytics process
CO4: Demonstrate the visualization and present the inference using various tools
CO5: Discuss the ethics surrounding privacy and algorithmic decision-making.

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 labor 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“Data Mining Concepts and Techniques”, Third Edition, Elsevier, 2011.

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

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

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

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%

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. Apply supervised learning techniques, including the Delta Rule and stochastic gradient descent (sgd), to train single-layer neural networks.

CO2. Build the multi-layer neural network using back-propagation algorithm and different learning rules.

CO3. Apply neural networks for binary and multiclass classification problems, as well as for regression problems.

CO4. Solve the data driven problems using deep neural networks, including vanishing gradients, overfitting, and computational load.

CO5. Design a full CNN architecture, incorporating strides and layers to condense information for image processing tasks.

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. MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence, Phil Kim, 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%

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. Apply the concept of big data and explore its business implications.

CO2. Explain the components of Hadoop and Hadoop Eco-System

CO3. Demonstrate the processing data on distributed file system

CO4. Develop big data solutions using hadoop eco system and apply machine learning techniques using R

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:

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

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.** Analyze portfolio characteristics using single and multi-factor models, including the application of style attribution analysis.
- **CO2.** Apply classical volatility and correlation models to estimate and forecast variance, covariance, and correlation matrices effectively.
- **CO3.** Evaluate time series models and their cointegration, assessing responses to shocks and integrating continuous time models.
- **CO4.** Develop advanced econometric techniques such as quantile regression for parameter estimation and inference.

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:

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

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. Analyze the components of data warehousing, including architecture, DBMS schemas, OLAP, and reporting tools.

CO2. Utilize the knowledge effectively to design and normalize databases, ensuring data integrity and organization in line with best practices.

CO3. Evaluate the optimize complex SQL queries using joins and subqueries for efficient data retrieval and manipulation

CO4. Create the skills to manage views in both centralized and distributed DBMSs while implementing robust data security measures.

CO5. Analyzing data mining concepts using multidimensional analysis and descriptive techniques to extract valuable insights

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%

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. Recall the requirements for the real world problems.

CO2. Demonstrate the project successfully by visualizing, analysing, testing and interpreting.

CO3. Construct data analytical skills.

CO4. Examine the findings of the study conducted in the preferred domain

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%

CO-PO Correlation Matrix for the Programme
B. Sc (Statistics and Data Analytics)

Course Code	Course Name	COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
SDS101	**Probability Theory**	**CO1**	3	2	2	1	2	1	1	1	1	1	1	1	2	3	2
		CO2	3	3	2	2	2	2	1	2	1	1	1	2	3	3	2
		CO3	3	3	3	2	3	3	2	2	1	1	1	2	3	3	3
		CO4	3	3	3	3	3	3	2	2	1	1	1	2	3	3	3
			3	**2.75**	**2.5**	**2**	**2.5**	**2.25**	**1.5**	**1.75**	**1**	**1**	**1**	**1.75**	**2.75**	**3**	**2.5**
SDS102	**Descriptive Statistics**	**CO1**	3	3	3	3	3	2	2	2	1	2	2	2	3	3	3
		CO2	3	3	3	3	3	2	2	2	1	2	2	2	3	3	3
		CO3	3	3	3	3	3	2	2	2	1	2	2	2	3	3	3
		CO4	3	3	3	3	3	2	2	2	1	2	2	2	3	3	3
			3	**3**	**3**	**3**	**3**	**2**	**2**	**2**	**1**	**2**	**2**	**2**	**3**	**3**	**3**
SDS103	**Probability Distribution**	**CO1**	3	2	2	2	2	1	1	1	1	1	1	1	2	2	2
		CO2	3	3	3	3	3	2	2	2	2	2	2	3	3	3	3
		CO3	3	3	3	3	3	2	2	2	1	1	1	3	3	3	3
		CO4	2	3	3	3	3	2	3	2	2	2	3	3	3	3	3
			2.75	**2.75**	**2.75**	**2.75**	**2.75**	**1.75**	**2**	**1.75**	**1.5**	**1.5**	**1.75**	**2.5**	**2.75**	**2.75**	**2.75**
SDS104	**Sampling Theory**	**CO1**	3	3	3	3	2	2	2	1	1	2	2	2	3	3	2
		CO2	3	3	3	3	3	2	2	1	1	2	2	2	3	3	2
		CO3	3	3	3	3	3	2	2	1	1	2	2	2	3	3	2
		CO4	3	3	3	3	3	2	2	1	1	2	2	2	3	3	2
			3	**3**	**3**	**3**	**2.75**	**2**	**2**	**1**	**1**	**2**	**2**	**2**	**3**	**3**	**2**
SDS201	**Statistical Inference**	**CO1**	3	2	3	-	3	2	-	-	-	-	-	2	2	2	3
		CO2	-	-	3	3	2	-	2	-	-	-	-	2	3	2	2
		CO3	-	3	3	-	2	2	-	-	-	-	-	-	2	3	3
		CO4	-	-	-	-	3	3	-	-	-	-	2	2	3	3	3
			3	**2.5**	**3**	**3**	**2.5**	**2.33**	**2**	**-**	**-**	**-**	**2**	**2**	**2.5**	**2.5**	**2.75**

SDS202	**Linear Algebra and Multivariate Analysis**	**CO1**	3	2	3	2	3	2	1	1	1	1	1	2	3	3	3
		CO2	3	3	2	2	3	2	2	2	2	2	2	3	3	3	3
		CO3	2	3	3	3	3	2	2	2	1	2	2	3	3	3	3
		CO4	2	3	3	3	3	3	2	2	1	2	3	3	3	3	3
			2.5	**2.75**	**2.75**	**2.5**	**3**	**2.25**	**1.75**	**1.75**	**1.25**	**1.75**	**2**	**2.75**	**3**	**3**	**3**
SDS203	**Exploratory Data Analysis and Visualization**	**CO1**	3	2	2	2	2	2	2	1	1	1	2	2	3	2	2
		CO2	3	2	3	2	3	3	2	1	1	1	2	2	3	2	2
		CO3	3	3	3	3	3	3	2	1	1	1	2	3	3	3	3
		CO4	3	2	3	3	3	3	2	2	1	2	2	2	3	2	2
		CO5	3	2	2	2	2	2	1	1	1	1	2	2	2	2	2
		CO6	3	3	3	3	3	3	2	1	1	2	2	3	3	2	3
			3	**2.3**	**2.7**	**2.5**	**2.7**	**2.7**	**1.8**	**1.2**	**1**	**1.3**	**2**	**2.3**	**2.8**	**2.2**	**2.3**
MTH204	**Introduction To Linear Programming and Game Theory**	**CO1**	3	3	-	3	-	-	-	-	-	-	-	3	3	-	-
		CO2	3	3	3	3	2	-	-	-	-	-	-	-	3	3	3
		CO3	3	3	3	3	-	-	-	-	-	-	-	3	-	3	-
		CO4	3	-	3	3	-	-	-	-	-	-	-	-	-	3	-
			3	**3**	**3**	**3**	**2**	**-**	**-**	**-**	**-**	**-**	**-**	**3**	**3**	**3**	**3**
SDS205	**Data Analytics Principles and Techniques**	**CO1**	1	2	2	2	2	2	3	3	3	2	2	2	2	2	2
		CO2	2	2	1	2	1	2	3	3	3	2	2	2	2	2	2
		CO3	3	3	1	2	2	1	3	3	3	2	2	2	2	2	1
		CO4	3	3	2	1	2	2	3	3	3	2	2	2	2	2	2
		CO5	3	3	3	3	3	3	1	3	1	2	2	2	3	2	2
			2.4	**2.6**	**1.8**	**2**	**2**	**2**	**2.6**	**3**	**2.6**	**2**	**2**	**2**	**2.2**	**2**	**1.8**
SDS301	**Linear Models**	**CO1**	2	2	1	2	2	1	3	3	3	2	2	2	2	2	2
		CO2	2	3	1	2	1	2	3	3	3	2	2	2	2	2	2
		CO3	2	3	1	1	2	1	3	3	3	2	2	2	2	2	2
		CO4	2	3	1	1	2	1	3	3	3	2	2	2	2	2	2
			2	**2.75**	**1**	**1.5**	**1.75**	**1.25**	**3**	**3**	**3**	**2**	**2**	**2**	**2**	**2**	**2**

SDS302	**Elementary Data Science**	**CO1**	3	2	3	2	3	3	2	1	1	-	-	1	3	2	3
		CO2	3	2	3	3	3	3	2	1	1	-	-	1	3	3	3
		CO3	3	2	3	3	3	3	2	1	1	-	-	1	3	3	3
		CO4	3	2	3	3	3	3	2	2	3	1	1	2	3	2	3
			3	**2**	**3**	**2.75**	**3**	**3**	**2**	**1.25**	**1.5**	**1**	**1**	**1.25**	**3**	**2.5**	**3**
SDS303	**Index Number and Time Series analysis**	**CO1**	3	2	2	2	3	2	3	2	1	2	2	3	3	3	2
		CO2	3	3	3	3	3	3	2	2	1	3	2	3	3	3	3
		CO3	3	3	3	3	3	3	2	2	1	2	2	3	3	3	3
		CO4	3	3	3	3	3	3	2	2	1	3	2	3	3	3	3
			3	**2.75**	**2.75**	**2.75**	**3**	**2.75**	**2.25**	**2**	**1**	**2.5**	**2**	**3**	**3**	**3**	**2.75**
SDS350	**Internship (Summer Project)**	**CO1**	3	2	2	2	2	2	2	1	2	2	3	2	3	2	3
		CO2	2	3	3	3	3	3	2	2	2	3	3	3	3	3	3
		CO3	2	2	3	3	2	2	2	1	3	3	3	2	3	3	3
			2.3	**2.3**	**2.7**	**2.7**	**2.3**	**2.3**	**2**	**1.3**	**2.3**	**2.7**	**3**	**2.3**	**3**		
SDS304	**Survey Sampling**	**CO1**	3	3	3	3	3	2	3	2	2	3	3	3	3	3	3
		CO2	2	2	3	3	3	3	2	2	2	3	2	2	3	3	3
		CO3	2	2	3	3	3	3	2	2	2	3	2	2	3	3	3
		CO4	2	2	3	3	3	3	2	2	2	3	2	2	3	3	3
			2.25	**2.25**	**3**	**3**	**3**	**2.75**	**2.25**	**2**	**2**	**3**	**2.25**	**2.25**	**3**	**3**	**3**
SDS305	**Applied machine learning**	**CO1**	3	2	2	2	2	3	2	2	2	2	2	3	3	3	3
		CO2	2	3	3	3	3	3	2	2	2	3	2	3	3	3	3
		CO3	2	2	3	3	3	3	2	2	2	3	2	3	3	3	3
		CO4	3	3	3	3	3	3	3	2	3	3	2	3	3	3	3
			2.5	**2.5**	**2.75**	**2.75**	**2.75**	**3**	**2.25**	**2**	**2.25**	**2.75**	**2**	**3**	**3**	**3**	**3**

SDS306	**Database Management System**	**CO1**	3	2	2	1	1	3	2	2	1	2	2	3	3	3	3
		CO2	2	3	2	2	2	3	1	1	1	2	2	2	3	3	3
		CO3	3	2	3	2	2	3	1	1	1	2	2	3	3	3	3
		CO4	3	2	3	3	3	3	2	1	1	3	2	3	3	3	3
		CO5	3	2	2	2	3	3	1	1	1	2	2	3	3	3	3
			2.8	**2.2**	**2.4**	**2**	**2.2**	**3**	**1.4**	**1.2**	**1**	**2.2**	**2**	**2.8**	**3**	**3**	**3**
SDS307	**Quantitative Methods for Finance**	**CO1**	3	2	2	1	3	2	2	1	1	1	2	3	3	2	3
		CO2	3	2	3	2	3	3	2	1	1	2	2	3	3	3	3
		CO3	3	2	3	3	3	3	2	2	1	2	2	3	3	3	3
		CO4	3	2	2	3	3	3	2	2	1	3	2	3	3	3	3
		CO5	3	1	2	2	3	3	2	1	1	1	2	3	3	2	3
			3	**1.8**	**2.4**	**2.2**	**3**	**2.8**	**2**	**1.4**	**1**	**1.8**	**2**	**3**	**3**	**2.6**	**3**
SDS351	**Project Work**	**CO1**	3	2	2	2	2	2	2	1	1	2	2	3	3	2	3
		CO2	3	3	3	3	3	3	2	2	2	3	3	3	3	3	3
		CO3	3	3	3	3	3	3	3	2	2	2	2	3	3	3	3
		CO4	3	2	3	3	3	3	3	3	1	2	2	3	3	3	3
			3	**2.5**	**2.75**	**2.75**	**2.75**	**2.75**	**2.5**	**2**	**1.5**	**2.25**	**2.25**	**3**	**3**	**2.75**	**3**
SDS401	**Stochastic Processes & Queueing Theory**	**CO1**	3	2	2	2	2	3	1	1	1	1	1	2	3	2	3
		CO2	3	2	2	2	2	3	1	1	1	1	1	2	3	2	3
		CO3	3	2	3	3	3	3	2	2	1	2	1	2	3	2	3
		CO4	3	2	3	3	3	3	2	2	1	2	1	2	3	2	3
			3	**2**	**2.5**	**2.5**	**2.5**	**3**	**1.5**	**1.5**	**1**	**1.5**	**1**	**2**	**3**	**2**	**3**
SDS402	**Advanced Machine Learning and Applications**	**CO1**	3	3	3	3	2	3	2	2	1	2	2	3	3	3	3
		CO2	3	3	3	3	3	3	2	2	1	2	2	3	3	3	3
		CO3	3	3	3	3	3	3	2	2	1	2	2	3	3	3	3
		CO4	3	3	3	3	3	3	3	3	1	3	3	3	3	3	3
		CO5	3	3	3	3	3	3	3	3	2	3	3	3	3	3	3
			3	**3**	**3**	**3**	**2.8**	**3**	**2.4**	**2.4**	**1.2**	**2.4**	**2.4**	**3**	**3**	**3**	**3**

SDS403	**Pricing, Hedging and Optimization**	**CO1**	2	1	2	1	2	1	2	1	1	1	1	1	3	2	2
		CO2	1	1	3	1	2	1	2	1	1	1	1	1	3	2	2
		CO3	2	1	3	2	2	1	1	1	1	1	1	1	3	2	2
		CO4	2	1	2	2	2	1	1	1	1	1	1	1	2	2	2
			1.75	**1**	**2.5**	**1.5**	**2**	**1**	**1.5**	**1**	**1**	**1**	**1**	**1**	**2.75**	**2**	**2**
SDS404	**Design of Experiments**	**CO1**	3	3	3	3	-	-	-	-	-	-	-	-	3	3	3
		CO2	3	2	2	2	-	-	-	-	-	-	-	-	2	2	2
		CO3	3	2	2	3	-	-	-	-	-	-	-	-	3	2	2
		CO4	2	3	3	3	-	-	-	-	-	-	-	-	3	2	2
			2.75	**2.5**	**2.5**	**2.75**	-	-	-	-	-	-	-	-	**2.75**	**2.25**	**2.25**
MTH450	**Research Methodology**	**CO1**	3	3	3	2	2	2	2	2	2	2	2	3	3	2	3
		CO2	3	3	3	3	2	2	2	2	3	3	3	3	3	3	3
		CO3	3	2	3	3	2	3	3	2	3	2	2	3	2	2	3
		CO4	3	2	2	2	2	3	2	2	2	2	3	3	3	2	3
			3	**2.5**	**2.75**	**2.5**	**2**	**2.5**	**2.25**	**2**	**2.5**	**2.25**	**2.5**	**3**	**2.75**	**2.25**	**3**
SDS451	**Statistical Quality Control**	**CO1**	3	3	3	3	3	3	2	2	2	3	2	3	3	3	3
		CO2	2	2	3	3	3	3	2	1	1	2	1	3	3	3	3
		CO3	3	3	3	3	3	3	2	2	2	3	2	3	3	3	3
		CO4	2	2	3	3	3	3	2	2	1	2	2	3	3	3	3
			2.5	**2.5**	**3**	**3**	**3**	**3**	**2**	**1.75**	**1.5**	**2.5**	**1.75**	**3**	**3**	**3**	**3**
SDS452	**Data Analytics for Finance**	**CO1**	3	3	2	2	2	2	2	2	1	1	1	2	3	3	2
		CO2	3	2	3	3	3	3	2	1	1	2	2	2	3	3	3
		CO3	3	2	3	3	3	3	2	1	1	2	2	3	3	3	3
		CO4	3	2	3	2	2	3	2	1	1	2	3	2	3	2	2
		CO5	2	2	2	2	2	2	2	2	3	2	2	2	2	2	2
			2.8	**2.2**	**2.6**	**2.4**	**2.4**	**2.6**	**2**	**1.4**	**1.4**	**1.8**	**2**	**2.2**	**2.8**	**2.6**	**2.4**

SDS405	**Artificial Neural Network**	**CO1**	3	2	3	2	3	3	1	1	1	2	2	3	3	3	3
		CO2	3	3	3	3	3	3	2	1	1	2	2	3	3	3	3
		CO3	3	3	3	3	3	3	2	2	2	2	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	3	2	2	2	2	2	3	3	3	3
			3	**2.6**	**3**	**2.8**	**3**	**3**	**1.8**	**1.6**	**1.6**	**2**	**2**	**3**	**3**	**3**	**3**
SDS453	**Data Engineering**	**CO1**	2	3	2	2	2	1	2	3	3	2	2	2	2	2	2
		CO2	2	3	1	2	2	1	2	3	3	2	2	2	2	2	2
		CO3	3	3	1	2	2	1	3	3	3	2	2	2	2	2	2
		CO4	3	3	1	1	1	1	2	3	3	2	2	2	1	2	1
			2.5	**3**	**1.25**	**1.75**	**1.75**	**1**	**2.25**	**3**	**3**	**2**	**2**	**2**	**1.75**	**2**	**1.75**
SDS454	**Practical Financial Econometric**	**CO1**	3	2	3	2	3	3	2	2	1	2	2	3	3	2	3
		CO2	3	2	3	3	3	3	2	2	1	2	2	3	3	2	3
		CO3	3	2	3	3	3	3	2	2	2	3	2	3	3	3	3
		CO4	3	3	3	3	3	3	2	2	2	2	3	3	3	3	3
			3	**2.25**	**3**	**2.75**	**3**	**3**	**2**	**2**	**1.5**	**2.25**	**2.25**	**3**	**3**	**2.5**	**3**
SDS455	**Data Management and Data Warehousing**	**CO1**	2	1	3	2	2	2	1	1	1	2	1	1	2	1	2
		CO2	3	2	3	2	2	2	1	1	1	2	2	2	2	2	2
		CO3	2	2	3	2	2	2	1	1	1	2	2	1	2	2	2
		CO4	2	1	2	3	2	2	1	1	1	2	2	2	2	1	2
		CO5	3	2	2	3	3	2	2	2	1	2	2	2	3	2	3
			2.4	**1.6**	**2.6**	**2.4**	**2.2**	**2**	**1.2**	**1.2**	**1**	**2**	**1.8**	**1.6**	**2.2**	**1.6**	**2.2**
SDS499	**Project/ Dissertation**	**CO1**	2	2	2	2	1	1	1	1	1	1	1	1	2	1	2
		CO2	3	2	3	3	2	2	1	1	1	2	2	2	3	2	2
		CO3	2	2	3	2	2	2	1	1	1	2	2	1	2	1	2
		CO4	2	1	2	3	2	2	1	1	1	2	2	2	2	1	2
			2.25	**1.75**	**2.5**	**2.5**	**1.75**	**1.75**	**1**	**1**	**1**	**1.75**	**1.75**	**1.5**	**2.25**	**1.25**	**2**