DEPARTMENT OF STATISTICS
UNIVERSITY COLLEGE OF SCIENCE
OSMANIA UNIVERSITY, HYDERABAD – 500 007

M.Sc. (STATISTICS) III-SEMESTER

SCHEME OF INSTRUCTIONS AND EXAMINATION FOR 2022-2024 BATCH

Paper	Sub. Code	Paper Title	Credits	Instruction Hours per Week	Semester end Exam duration	Max. Marks in Semester end Exam	Max. Marks in the Internal Assessment and Assignments
THEORY PAPERS							
I	STS-301	Non-Parametric Inference (NPI)	3	3	3	70	20+10
II	STS-302	Quality control & Optimization Techniques (QCOT)	3	3	3	70	20+10
III	STS-303 **Elective – I**	A) Applied Regression Models (ARM) B) Econometric Models (EM) C) Advanced Design & Analysis of Experiments (ADAE). D) Statistical Process in Data Science	3	3	3	70	20+10
IV	STS-304 **Elective – II**	A) Data Mining Techniques (DM) B) Bayesian Inference (BI) C) Advanced Machine Learning Techniques (MDLT)	3	3	3	70	20+10
PRACTICALS PAPERS							
V	STS-305	Non-Parametric Inference & Optimization Techniques (NPI & QCOT)	2	4	2	50	-
VI	STS-306	Practicals on E-I & E-II	2	4	2	50	-
VII	STS-307	Statistical Analysis using R & TORA	2	4	2	50	-
VIII	STS-308	Data Analysis Project (Mini project)	2	4	2	**50**	-
Semester Total			**20**	**12+16***	-	**600**	

M. SC. (STATISTICS) III-SEMESTER

STS-301: PAPER I - NON PARAMETRIC INFERENCE (NPI)

Unit–I

Non-parametric density estimation: Density estimates, survey of existing methods. Rosenblatt's naïve density estimator, its bias and variance. Consistency of Kernel density estimators and its MSE.

Unit–II

Nonparametric Tests: one-sample problems based on sign test, Wilcoxon signed Rank test, run test and Kolmogorov – Smirnov test. Two sample problems based on sign test, Wilcoxon signed rank test for paired comparisons, Wilcoxon Mann-Whitney test, (Expectations and variances of above test statistics, Statements about their exact and asymptotic distributions).

Unit–III

Nonparametric Tests: Two sample problems based on Kolmogorov – Smirnov Test, Wald–Wolfowitz Runs test and Normal scores test. Ansari–Bradley test for two sample dispersions. (Expectations and variances of above test statistics, Statements about their exact and asymptotic distributions).

Unit–IV

Nonparametric Tests: Chi–Square test of goodness of fit and independence in contingency tables. Tests for independence based on Spearman's rank correlation and Kendall's Tau. Kruskal–Wallis test for one-way layout (K-samples). Friedman test for two-way layout (randomised block).

Unit–V

Asymptotic Relative Efficiency (ARE) and Pitman's theorem. ARE of one sample, paired sample and two sample locations tests. The concept of Rao's second order efficiency and Hodges–Lehman's deficiency with examples.

REFERENCES

1. Gibbons – Non-parametric Statistical Inference (1978)
2. Myles Hollander and Douglas A. Wolfe: Nonparametric statistical methods (John Wiley)
3. Silverman: Density estimation for statistics and data analyses.
4. W.J. Conover – Practical Non parametric Statistics (John Wiley)
5. Sidney Siegel – Non-parametric Statistics for Behavioural Science, Mc. Graw Hill.
6. Ferguson, T.S. – Mathematical Statistics, A decision theoretic approach (Academic press)

M.SC. (STATISTICS) III-SEMESTER

STS-302: PAPER II: QUALITY CONTROL AND OPTIMIZATION TECHNIQUES

Unit–I

Review of control charts for variable data and attributes : O.C. and A.R.L. functions of control charts for variables and attributes, modified control charts for variables and Acceptance control charts for attributes, control by gauging. Moving Average and exponentially weighted moving average charts, Cu-sum charts using V-Masks and decision intervals.

Unit–II

Process Capability Analysis: Capability indices Cp, Cpk and Cpm, estimation, confidence intervals and tests of hypotheses relating to capability indices for normally distributed characteristics. Acceptance sampling plans for attributes, single, double and sequential sampling plans and their properties

Unit-III

Rectifying inspection plans for attributes, AOQ, AOQL, designing of Rectifying Sampling Plans for specified AOQL and LTPD. Sampling Plans for inspection by variables for one–sided and two–sided specifications; Dodges Continuous sampling Plan–l and its properties, modifications over CSP–l.

Unit–III

Review on LPP, Graphical & simplex, Charmers methods, Duality in LPP; Duality and Complementary slackness theorems. Primal and dual relation. Dual simplex Algorithm; Sensitivity Analysis: Discrete changes requirement and cost vectors; parametric programming: Parameterisation of cost and requirement vectors.

Integer Programming Problem: Gomory's cutting plane Algorithm for pure and mixed IPP Branch and bound Technique.

Unit–IV

Basic concepts of Networks constraints; Construction of Network and critical path; PERT and CPM; Network flow problems. Time Cost Analysis.

Inventory: Introduction; ABC analysis and Deterministic Inventory models with and without shortages.

REFERENCES

1. Montgomery, D.C.(1985) : Introduction to Statistical Quality Control, Wiley
2. Wetherill, G.B. (1977): Sampling Inspection and Quality Control, Halsted Press.
3. Cowden, D. J. (1960) : Statistical Methods in Quality Control, Asia Publishing House.
4. Kantiswarup; Gupta P.K. and Singh, M.N.(1985) : Operations Research; Sultan Chand
5. Taha, H.A.(1982): Operations Research : An Introduction; MacMillan
6. Sharma,S.D.: Operations Research.
7. Ott,E.R. (1975) : Process Quality Control, McGraw Hill
8. Phadke, M.S. (1989): Quality Engineering through Robust Design, Prentice Hall.
9. Wetherill, G.B., and Brown, D.W: Statistical Process Control: Theory and Practice, Chapman and Hall.
10. Hillier F.S. and Leiberman,G.J.(1962) : Introduction to Operations Research; Holdon Day

M.SC. (STATISTICS) III-SEMESTER

ELECTIVE-I: STS-303(A): PAPER-III (A): APPLIED REGRESSION MODELS (ARM)

Unit–I

Selection of best linear regression: Introduction to selection of best linear regression, all possible regression, backward, forward, step-wise, stage-wise regressions. Ridge regression.

Unit–II

Non-linear regression: Introduction to non-linear regression model, some commonly used families of non-linear regression functions, statistical assumptions and inferences for non-linear regression, linearizable models, determining the Least squares estimates, The Gauss – Newton method, ML estimation, (D&S).

Unit–III

Logistic regression model: Introduction to simple Logistic model, Fitting the model, testing for the significance of the coefficients, Logistic model for Dichotomous independent variable; Introduction to multiple Logistic regression, fitting the multiple logistic regression model, testing for the significance of the model.

Unit–IV

Probit Analysis: Introduction, Analysis of Biological data, sigmoid curve, fitting a Probit Regression line through least squares method.

Robust Regression: Introduction, Least absolute deviations regression (L_1 Regression), M–estimators, examples, and Least Median of Squares (LMS) regression, Robust Regression with Ranked Residuals.

Unit-V

Generalized Linear Models: Introduction, the exponential family of distributions, fitting GLIM. Concept of Mixed, Random Effects and Fixed Models–Introduction, General description, estimation, estimating variance components from balanced data.

REFERENCES

1. Regression Analysis: Concepts and Applications, Franklin A. Graybill and Hariharan K. Iyer
2. Applied Regression Analysis: Norman R. Draper and Harry Smith
3. Applied Regression Analysis, linear models and related methods: John Fox
4. Non–linear Regression Analysis and its Applications: Douglas M. Bates and Donald G. Watts
5. Applied Logistic Regression: David W. Hosme and Stanley Lemeshow.
6. Linear Models for unbalanced Data: Shayler Searle
7. Residuals and Influence in Regression: R. Dennis Cook and Sanford Weisberg
8. Log–linear models and Logistic Regression: Ronald Christensen.

M.SC. (STATISTICS) III-SEMESTER

ELECTIVE-I: STS-303(B): PAPER-III (B): ECONOMETRIC MODELS (EM)

Unit–I

Meaning and scope of econometrics. Concepts of dummy variables and proxy variable. Problems and methods of estimation in single equation regression Models

Multicollinearity: Consequences of multicollinearity, tests to detect its presence and solutions to the problem of multicollinearity.

Unit–II

Generalised Least Squares: Estimates of regression parameters – Properties of these estimates.

Heteroscedasticity: Consequences of heteroscedastic disturbances – test to detect its presence and solutions to the problem of heteroscedasticity.

Unit–III

Auto Correlation: Consequences of autocorrelated disturbances, Durbin – Watson test – Estimation of autocorrelation coefficient (for a first order autoregressive scheme).

Distributed lag models: study of simple finite lag distribution models – Estimation of the coefficients of Kayak geometric lag model.

Instrumental Variable: Definition – derivation of instrument variable estimates and their properties.

Unit–IV

Errors in variables: Problem of errors in variables simple solutions using instrumental variables technique. Simulation equation models and methods of estimation: distinction between structure and Model–Exogenous and Endogenous variables – Reduced form of a model.

Unit–V

Problem of identification – Rank and order conditions and their application.

Methods of estimation: Indirect least squares. Two stages least squares, three stages least squares. A study of merits and demerits of these methods.

REFERENCES

1) Johnston – Econometrics Methods (2^{nd} Edition) :
2) G. S. Maddala – Econometrics
3) A. Koutsoyiennis – Theory of econometrics

M.SC. (STATISTICS) III-SEMESTER

ELECTIVE-I: STS-303(C): PAPER-III(C): ADVANCED DESIGN OF EXPERIMENTS (ADE)

Unit–I

General incomplete block designs and its information matrix. Balanced Incomplete block design (BIBD) – Parametric relations, intra–block analysis, recovery of inter–block information. Concepts of Symmetric, Resolvable and Affine resolvable BIBDS. Construction of BIBDS using MOLS.

Unit–II

Partially balanced incomplete block design with two–associate classes PBIBD(2)–Parametric relations, intra–block analysis, Four different association schemes.

UNIT-III

Youden Square design and its analysis. Lattice designs, Balanced Lattice Design, Simple Lattice Design and their analysis. Construction of Youden square, balanced lattice designs

Unit–IV

Concept of Response surface methodology (RSM), the method of Steepest ascent. Response surface designs–designs for fitting first–order and second– order models, Variance of estimated response. Second order rotatable designs (SORD), central composite designs (CCD)–role of CCD as alternative to 3^k designs, rotatability of CCD.

Unit–V

Experiments with mixtures–Simplex Lattice designs, first-order and second-order mixture models and analysis. Optimum designs–various optimality criteria and their interpretations. Repeated measurements designs. Cross–over designs and Row–Column designs.

REFERENCES

1. Montgomery, D.C.: Design and Analysis of Experiments
2. Parimal Mukhopadhyay : Applied Statistics
3. Das, M.N., and Giri, N.: Design and Analysis of Experiments
4. Myers, R.H. : Response Surface Methodology
5. Aloke Dey : Theory of Block Designs
6. Cornell, M : Mixture Experiments

M.SC.(STATISTICS) III-SEMESTER

ELECTIVE-I : STS-303 (D): PAPER-III (D) : STATISTICAL PROCESS IN DATA SCIENCE

Unit – I

Data Visualization: Data types, Measurement of scales, understanding data with descriptive statistics. Data visualization techniques: Pictogram, Pie Chart, Bar Chart, Histogram, Line plot, frequency curves& polygons, ogive curves, Scatter Plot, Gantt Chart, Heat Map, Box and Whisker Plot, Waterfall Chart, Area Chart, Stacked Bar Charts - Sub Plots – Matplotlib, Seaborn Styles, Box plot - Density Plot - Tree map - Graph Networks. Visual Perception and Cognition, Applications of Principles of Information Visualization, Dashboard Design.

Unit-II

Data Pre-processing: Understanding data with Descriptive statistics. Data pre-processing steps, Data transformations (Standardize, Normalize, converting data from one scale to other scales). Identification suitable basic statistical tools / tests Parametric tests (z-, χ^2, t-, F-tests), Nonparametric tests (Sign test, Median, Wilcoxon sign rank, Mann-Whitney U, K-S, Wald-Wolfowitz run test) for the data sets. Feature selection methods

Unit-III

Introduction to Data Modelling: Review of the modelling process, Concepts of Classification & Clustering, Supervised and Un-supervised Modelling, Concepts of Model evolution, Cross validation concepts, (train/test, K fold and Leave out one approaches), Model Performance evaluation for Qualitative and Quantitative data, Model improvement and saving models for future use (classification matrix, Precision and Recall, F1 score, Sensitivity, Specificity, ROC curve) and Model performance concepts for regression (MSE, RMSE, R^2, adj R^2, MAPE),

Unit-IV

Concepts of Model improvement (Tuning parameters using manual search, Manual grid search, random search) and saving models for future use. Simple linear regression and its analysis (model fitting, regression ANOVA, testing lack of fit, MSE, RMSE, R^2, adj R^2, testing regression coefficients and confidence limits).

Unit-V

Basic concepts on Multivariate data; Simple, Partial & Multiple correlations; Multi collinearity; Multiple linear regression and its analysis; Selection of best linear regression (over fitting & under fitting) & its methods in outline (all possible, forward, backward, step-wise and stagewise). Simple and Multiple Logistic models fitting and its analysis.

REFERENCES

1) Foster Provost & Tom Fawcett, Data science for Business, O'REILLY Publications
2) Henrik Brink, Joseph W. Richards. Mark Fetherolf, Real World Machine Learning, Manning Publications.
3) Foster Provost & Tom Fawcett, Data science for Business, O'REILLY Publications
4) Henrik Brink, Joseph W. Richards. Mark Fetherolf, Real World Machine Learning, Manning Publications
5) Brett Lantz, Machine Learning with R, Packt Publications.

M.SC.(STATISTICS) III-SEMESTER

ELECTIVE-II : STS-304(A): PAPER-IV(A) : DATA MINING (DM)

Unit-I

Introduction: Challenges, Origins of Data Mining, Data Mining Tasks; **Data:** Types of Data, Data Quality, Data Preprocessing, Measures of Similarity and Dissimilarity; **Exploring Data:** Visualization, OLAP and Multidimensional Data Analysis

Unit-II

Classification: Preliminaries, General approach to solving a classification problem, Decision tree induction, Model Over-fitting, – Evaluating the performance of a classifier – Methods of comparing classifiers; Rule-based classifier, Nearest-Neighbor classifiers, Bayesian classifiers

Unit-III

Classification: Artificial Neutral Networks, Perceptron classifier, Support vector machine, Ensemble methods, Class imbalance problem – Multiclass problem

Unit-IV

Cluster Analysis: Agglomerative hierarchical clustering, K-means, DBSCAN, C4.5, CART Cluster evaluation.

Unit-V

Association Analysis: Problem definition, Frequent item set generation, Rule generation, Compact representation of frequent item sets, Alternative methods for generating frequent item sets, FP-Growth Algorithm, Evaluation of Association patterns, Effect of Skewed support distribution; Handling categorical attributes. Handling continuous attributes, Handling a concept hierarchy.

REFERENCES

1. Pang-Ning Tan, Michael Steinbach, Vipin Kumar (2008): “Introduction to Data Mining”, Pearson Education.
2. Arun K Pujari, Data Mining Techniques, University Press, 2nd Edn, 2009.
3. K.P. Soman, Shyam Diwakar, V.Ajay, Insight into Data Mining Theory and Practice, PHI, 2010.
4. Vikram pudi P. Radha Krishna, Data Mining, Oxford University Press, 1st Edition 2009
5. Galit S, Nitin RP, Peter C Bruce. Data Mining for Business Intelligence. Wiley India Edition,2007.

M.SC. (STATISTICS) III-SEMESTER

ELECTIVE-II : STS-304(B): PAPER-IV(B): BAYESIAN INFERENCE (BI)

Unit-I

Bayes theorem, Bayesian Concept to priors and posteriors, computation of the posterior distribution.) subjective prior distribution, Conjugate family of priors of a parameter. Hyper parameters of a prior from conjugate family. Conjugate families for (i) exponential family models, (ii) models admitting sufficient statistics of fixed dimension.

Unit-II

Subjective prior distribution of a parameter. Non informative, improper and invariant priors. Jeffrey's invariant prior. Bayesian point estimation as a prediction problem from posterior distribution. Bayes estimators for (i) absolute error loss (ii) squared error loss (iii) 0 - 1 loss. Generalization to convex loss functions.

Unit-III

Evaluation of the estimate in terms of the posterior risk. Bayesian interval estimation: Credible intervals. Highest posterior density regions. Interpretation of the confidence coefficient of an interval and its comparison with the interpretation of the confidence coefficient for a classical confidence interval.

Unit-IV

Bayesian testing of Hypothesis: Specification of the appropriate form of the prior Distribution for a Bayesian testing of hypothesis problem. Prior odds, Posterior odds, Bayes factor for various types of testing hypothesis problems depending upon whether the null hypothesis and the alternative hypothesis are simple or composite. Bayesian prediction problem. Large sample approximations for the posterior distribution.

UNIT-V

Estimation of parameters using Markov Chain Monte Carlo methods: Gibbs Sampler and Metropolis-Hasting Method and other computer simulation methods. Bayesian calculations for non-conjugate priors: (i) Importance sampling, (ii) Obtaining a large sample of parameter values from the posterior distribution using Acceptance - Rejection methods.

REFERENCES

1. Berger, J. O. Statistical Decision Theory and Bayesian Analysis, Springer Verlag.
2. Robert C. P. and Casella, G. Monte Carlo Statistical Methods, Springer Verlag.
3. Leonard T. and Hsu, J. S. J. Bayesian Methods. Cambridge University Press.
4. Box, G. P. and Tiao, G. C. Bayesian Inference in Statistical Analysis, Addison - Wesley.

M.SC.(STATISTICS) III-SEMESTER

ELECTIVE-II : STS-304(C): PAPER-IV(C) : ADVANCED MACHINE LEARNING TECHNIQUES

Unit – I

Basic Concepts to Statistical Pattern Recognition, Pattern Recognition System, Fundamental problems in Pattern Recognition. Linear classifies, Multiple Linear regression, Logistic regression, Linear Discriminant Function (for binary outputs) with minimum squared error, Naïve Bayes classifier, Support Vector Machines, KNN algorithm

Unit – II

Decision Tree algorithms, Random Forest algorithm, Bagging, Gradient boosting, Ada-Boosting and XG-Boosting algorithm, Market-Basket Analysis.

Unit – III

Cluster Analysis: Introduction, similarities and dissimilarities, Hierarchical clustering, Single linkage method, k-means and k-Nearest Neighbourhood (KNN) clustering,

Unit – IV

Introduction to Artificial Neuron Networks and its characteristics; Algorithms of Perceptron Learning; Multi-layer Perceptron Learning, Gradient Descent Learning, Least Mean Square learning, Widrow-Hoff Learning. Back-Propagation and their applications.

UNIT – V

Reinforcement learning, Markov Decision Process, Hidden Markov Model, Convolutional Neural Networks, Recurrent Neural Networks, Long-Short Term Memory Networks.

REFERENCES

1. Shai Shalev-Shwartz, Shai Ben-David Understanding Machine Learning: From Theory to Algorithms, Cambridge University press.
2. Marc Peter Deisenroth, A Aldo faisal, Cheng soon Ong: “Mathematics for Machine Learning”, Cambridge University Press, First Edition.
3. Hayes: Artificial Neural networks

M.SC.(STATISTICS) III-SEMESTER

STS-305: PAPER-V : NON-PARAMETRIC INFERENCE, QUALITY CONTROL & OPTIMIZATION TECHNIQUES (NPI & QCOT)

PRACTICAL -I (CONVENTIONAL)

SECTION-A: LIST OF PRACTICALS ON NON–PARAMETRIC INFERENCE

1. Sign test and Wilcoxon signed rank test (including paired comparison)
2. Run test for randomness
3. Two Samples:
    a) Wilcoxon Mann-Whitney test
    b) Kolmogorov – Smirnov test
    c) Wald Wolfowitz test
4. Goodness of fit: Chi–square and Kolmogorov – Smironov test
5. Normal Scores test
6. Kruskal–Wallis for one–way layout
7. Friedman test for two–way layout
8. Tests for independence in contingency tables: Spearman's rank correlation, Kendall's Tau
9. Ansari-Bradley test for two sample dispersions.

SECTION-B: LIST OF PRACTICALS ON QUALITY CONTROL, OPTIMIZATION TECHNIQUES

1. Construction of OC and ASN curves for $\overline{X}$ and R - charts
2. Construction of OC and ASN curves for Control charts for attributes (p,d,c; for a fixed n)
3. Construction of simple and exponentially weighted moving average control chart and simple moving range control chart.
4. Construction of CUSUM charts V – Mark and Tabular methods
5. Construction of AOQ and AFI curves for CSP–I
6. Dual Simplex Method
7. Sensitivity Analysis
8. Parametric Programming Problem
9. Integer Programming Problem
10. Evaluation of project time through CPM and PERT
11. Evaluation of Time cost analysis through CPM and PERT

M.SC. (STATISTICS) III-SEMESTER

STS-306: PAPER-V: SECTION-A (ELECTIVE-I) & SECTION-B (ELECTIVE-II)

PRACTICAL-II (CONVENTIONAL & USING SOFTWARE)

ELECTIVE-I (A): LIST OF PRACTICALS APPLIED REGRESSION MODELS

1. Problems on All possible Regression using R^2.
2. Problems on Stage wise Regression.
3. Computation of odds ratio (Dichotomous).
4. Computation of Multiple Logistic regression.
5. Fitting a probit regression line through least squares method.
6. Computation of variance components.
7. Computation of mean and variance for exponential family of distributions.

ELECTIVE-I (B): LIST OF PRACTICALS ON ECONOMETRIC MODELS

1. Use of dummy variables (dummy variable trap) and seasonal adjustment
2. GLS estimation and predictors
3. Tests for heteroscedasticity.
4. Tests for Autocorrelations
5. Instruments variable estimation
6. Estimation with lagged dependent variable
7. Identification problems – Checking rank and order condition
8. Two SLS estimation

ELECTIVE-I (C): LIST OF PRACTICALS ON ADVANCED DESIGNS AND ANALYSIS OF EXPERIMENTS

1. Intra-block analysis of BIBD
2. Analysis of Youden Square Design
3. Intra-block analysis of PBIBD (2)
4. Analysis of Balanced Lattice design
5. Analysis of Simple Lattice design
6. Fitting of Response surface design model for first and second order
7. Construction of rotatable design
8. Analysis of Mixture Experiments.

ELECTIVE-I(D): STATISTICAL PROCESS IN DATA SCIENCE USING PYTHON

Data sets of Kaggle.com can be used for practice. For example, few of the them are: Iris Dataset; flights.csv Dataset; Sustainable Development Data; Credit Card Fraud Detection; Employee dataset; Heart Attack Analysis & Prediction Dataset; Dataset for Facial recognition; Covid_w/wo_Pneumonia Chest Xray Dataset; Groceries dataset; Financial Fraud and Non-Fraud News Classification; IBM Transactions for Anti Money Laundering

List of Practicals in Data Handling with Python (USING Packages):

1. Understanding data with Data types, Measurement of scales, descriptive statistics and data pre-processing steps.
2. Data transformations (Standardize, Normalize, converting data from one scale to other scales).
3. Parametric tests (z-, χ^2, t-, F-tests, ANOVA), Correlation & Regression etc.
4. Non-Parametric tests (Sign test, Median, Wilcoxon sign rank, Mann-Whitney U, Run test).
5. Applying the modelling process, Model evolution, over fitting, under fitting, cross validation concepts, (train/test, K fold and Leave out one approaches),
6. Evaluation of Model Performance for classification techniques for qualitative and Quantitative data.
7. Drawing One dimensional diagrams (Pictogram, Pie Chart, Bar Chart,).
8. Drawing two-dimensional (Histogram, Line plot, frequency curves & polygons, ogive curves, Scatter Plot)
9. Drawing Gantt Chart, Heat Map, Box - Whisker Plot, Correlation Matrices.

ELECTIVE-II (A): LIST OF PRACTICALS IN DATA MINING

1. Nearest-Neighbor classifiers
2. Bayesian classifiers
3. Support vector machine K-means
4. DBSCAN
5. Compact representation of frequent item sets
6. FP-Growth Algorithm

ELECTIVE-II (B): LIST OF PRACTICALS IN BAYESIAN INFERENCE.

1. Data Simulation for Uniform, Normal, Exponential, Cauchy and Poisson Distributions.
2. Bayesian estimation of parameters for p in Binomial(n,p) with their conjugate paired distributions using Metropolis Hasting and Gibbs Sampler).
3. Bayesian Estimation of parameters and μ in Normal (μ, σ^2) distribution with their conjugate paired distributions (using R) with Metropolis Hasting / Gibbs sampler.

ELECTIVE-II(C): LIST OF PRACTICALS ON ADVANCED MACHINE LEARNING TECHNIQUES

(Implementation using Python)

1. Multiple linear regression and Multiple Logistic regression
2. KNN & K-means
3. Naïve Bayes classier
4. Support vector machines
5. Random forest
6. Bagging and Boosting
7. Implementation of Perceptron Learning Algorithm.
8. Implementation of Multi-layer Perceptron Learning
9. Implementation of Back-Propagation Algorithms.
10. Implementation of Hidden Markov Model,

M.SC. (STATISTICS) III- SEMESTER

STS-307: PAPER VII: STATISTICAL ANALYSIS USING R AND TORA
PRACTICAL-III

Practical with R Package for the following topics.

1. **Data Visualization**: Pie diagram, Bar diagram, Histogram, Line plot, frequency curves & polygons, Scatter Plot, Gantt Chart, Box Plot.
2. **Descriptive Statistics:** Measures of Central Tendencies, Dispersions, Relative measures of Dispersions, Moments, Skewness, Kurtosis.
3. **Parametric Tests**: Testing for Mean(s), Variance(s), Proportion(s), ANOVA for one-way two-way and two way with one and m-observations per cell and with & without interactions,
4. **Non–Parametric tests:** Sign test, Wilxon Sign Rank test, Mann-Whitney U-test, Run test, Kolmogorov Smirnov test, Chi-square test for goodness of fit and Chi-square test independence.
5. **Design & Analysis of Experiments:** Analysis of Variances for Completely randomized, randomized block and latin Square Designs and Factorial experiments (2^2, 2^3 F.E. without confounding).
6. **Regression Analysis:** Analysis of Simple and Multiple Linear Regression models, Selection Best Linear Regression Model (All possible, forward, backward, stepwise and stage wise methods). Binary and multinomial Logistic regression models, Probit analysis.
7. **Multivariate Data Analysis:** Linear Discriminant Analysis, Principal Component analysis, Factor analysis, Multi-dimensional scaling, Cluster analysis.
8. **Statistical Quality Control:** Construction Control charts for variables and attributes.

TORA

Operations Research (TORA Package):

Solving a Linear Programming Problems: Graphical method, simplex method, Big-M method, two Phase method, Duality, Dual simplex, transportation problem, Assignment Problem, sensitivity analysis.

M.SC.(STATISTICS) III- SEMESTER

STS-308: PAPER VII: DATA ANALYSIS MINI PROJECT
PRACTICAL-IV

Objectives and Outcomes:

1. To familiarize tools and techniques and content for presentation
2. To enhance practical presentation, effective communication and professional skills
3. To expose the students to answer the queries raised on the topic of presentation.
4. To encourage students to work with innovative and entrepreneurial ideas
5. Demonstrate the ability to synthesize and apply the knowledge and skills acquired in the academic program to real-world problems
6. Evaluate different solutions based on economic and technical feasibility
7. Effectively plan a project and confidently perform all aspects of project management
8. Demonstrate effective written and oral communication skills

PROJECT GUIDELINES:

1. The Head of Department will appoint Internal supervisor to Guide the students in each group.
2. Each group should consist of Five students.
3. Each student in the group must actively participate and report to the internal supervisor.
4. Each group has to search for the internship from any industry/ institution, if not found they have to choose a project with the help of supervisor allotted such that, the aim of project work is to develop solutions to realistic problems applying the knowledge and skills obtained on the courses studied with specializations, new technologies and current industry practices.
5. Each student has to give minimum two seminars, one in the second week ("Project Design Seminar") another on 8^{th} week (project progress seminar).
6. Submit Title of the project and one page abstract /synopsis about the project in the first week to the Head, forwarded by the internal supervisor.
7. Each project should give a 30 minutes presentation using power point presentation and followed by 10 minutes of discussion.
8. Project seminar presentations should contain, source of the data, Sample data, data description, literature survey on the similar studies, objectives of the study, Methodology, statistical techniques, work plan etc. and details of progress of the work, individual roles and their work distribution and their plan etc.
9. Each group Project Report should follow the Ph.D. thesis norms with Plagiarism report and each group has to submit two copies duly signed by the Students, Supervisor, industry certificate (if exists) and Head of the Department on before the last instruct date of the semester.
10. Project Marks will be awarded based on all stages of the project and the topic chosen, seminar presentation, communication skills, role/ contribution of the student in the project etc and viva-voce conducted by the internal & External examiners.

M.SC. (STATISTICS) SEMESTER-IV

Paper	Sub. Code	Paper Title	Credits	Instruction Hours per Week	Semester end Exam duration	Max. Marks in Semester end Exam	Max. Marks in Internal Assessment and Assignments
THEORY PAPERS							
I	STS-401	Stochastic Processes (SP)	3	3	3	70	20+10
II	STS-402	Time Series Analysis (TSA)	3	3	3	70	20+10
III	STS-403	A) Advanced Operations Research (AOR) B) Text Analytics (TA) C) Demography (DGY)	3	3	3	70	20+10
IV	STS-404	A) Artificial Neural Networks (ANN) B) Design & Analysis of Algorithms (DAA) C) Clinical Trails (CT)	3	3	3	70	20+10
PRACTICAL PAPERS							
V	STS-405	Stochastic Processes & Time Series Analysis	2	4	2	50	-
VI	STS-406	Elective – I & II	2	4	2	50	-
VII	STS-407	Major Project	4	8	2	100	-
Total			20	12+16*	-	600	

M.SC.(STATISTICS) IV-SEMESTER
STS-401: PAPER-I: STOCHASTIC PROCESSES (SP)

UNIT – I

Introduction to stochastic processes; classification of stochastic process according to state-space and time-domain. Finite and countable state Markov chains; time-homogeneity; Chapman-Kolmogorov equations; marginal distribution and finite – dimensional distribution;.

UNIT – II

classification of states of a Markov chain – recurrent, positive recurrent, null - recurrent and transient states. Period of a state. Canonical form of transition probability matrix of a Markov chain. Fundamental matrix; probabilities of absorption from transient states into recurrent classes, in a finite Markov Chain; mean time for absorption. Ergodic state and ergodic chain.

Unit-III

Stationary distribution of a Markov chain. Existence and evaluation of stationary distribution. Random walk and gambler's ruin problem. Weiner process as limit of random walk. First passage time of the process.

UNIT – IV

Discrete state-space, continuous time Markov Processes – Kolmogorov difference - differential equations. Poisson process and its properties. Birth and Death Process, application in queuing. Pure Birth and pure Death processes.

UNIT – V

Renewal process, elementary renewal theorem and its applications. Statement and uses of Key – renewal theorem. Residual life time. Branching process – Galton-Watson branching process, mean and variance of size of n^{th} generation; probability of ultimate extinction of a branching process – fundamental theorem of Branching process – Examples.

REFERENCES

1. Medhi,J. (1982) : Stochastic Processes – Wiley Eastern
2. Karlin, S. and Taylor, H.M. (1975): A First Course in Stochastic Processes, Vol. I, Academic Press.
3. Bhat, B.R. (2000): Stochastic Models: Analysis and applications – New Age International India.
4. Basu, A.K. (2003): Introduction to Stochastic Process, Narosa Publishing House.

M.SC. (STATISTICS) IV-SEMESTER

STS-402: PAPER-II: TIME SERIES ANALYSIS (TSA)

Unit–I

Stationery stochastic processes. The autocovariance and Auto correlation functions and their estimation. Standard errors of autocorrelation estimates. Bartlett's approximation (without proof). The periodogram, the power spectrum and spectral density functions. Link between the sample spectrum and autocorrelation function.

Unit–II

Linear Stationary Models: Two equivalent forms for the general linear process. Autocovariance generating function and spectrum, stationarity and invertibility conditions for a linear process. Autoregressive and moving average processes, autocorrelation function (ACF), partial autocorrelation function (PACF).

Unit-III

Spectrum for AR processes up to 2. Moving average process, stationarity and Invertibility conditions. ACF and PACF for M.A. (q), spectrum for M.A. processes up to order 2. Duality between autoregressive and moving average processes, Mixed AR and MA(ARMA) process. Stationarity and invertibility properties. ACF and spectrum of mixed processes. The ARMA(1.1) process and its properties. Linear Non-Stationary Models – Autoregressive integrated and moving average (ARIMA) processes. The three explicit forms the ARIMA models (viz) Difference equation, random shock and inverted forms.

Unit–IV

Model Identification–Stages in the identification procedures. Use of autocorrelation and partial auto–correlation, functions in identification. Standard errors for estimated autocorrelation and partial autocorrelations. Initial estimates MA, AR and ARMA processes and residual variance.

Model Estimation: Least squares and Maximum likelihood estimation and interval estimation of parameters.

Unit–V

Model Diagnostic checking – checking the stochastic model diagnostic checks applied to residuals. Forecasting: Minimum mean square error forecasts and their properties, derivation of the minimum mean square error forecasts, calculating and updating forecasts at any lead time.

REFERENCES

1. Box and Jenkins: Time Series Analysis
2. Anderson, T.W. : Time Series Analysis.
3. Brockwell,P.J., and Davis,R.A.: Time Series : Theory and Methods (Second Edition). Springer–Verlag.

M.SC. (STATISTICS) IV-SEMESTER

ELECTIVE-I: STS-403(A): PAPER-III(A): ADVANCED OPERATIONS RESEARCH (AOR)

Unit–I

Non-linear Programming problem – Formulation, Generalised Lagrange multiplier technique, Kuhn-Tucker necessary and sufficient conditions for optimality of an NLPP, Wolfe's and Beale's Algorithms for solving QPP. Separable Programming Problem; Piecewise linear Approximation method. Linear Fractional Programming Problem and its applications.

Unit–II

Dynamic Programming, Principle of optimality, solution of LPP by Dynamic Programming technique, Knapsack problem by Dynamic Programming Technique. General goal Programming model and formulation of its objective function. Solutions to linear goal programming and linear integer goal programming.

Unit–III

Game Theory: Two person zero sum game, pure strategies with saddle point, mixed strategies with saddle point, principles of dominance and games without saddle point, 2xm, mx2, mx n games Decision Analysis: Introduction, Steps in Decision theory approach, Types of Decision making environments, Decision making under uncertainty – criterion of optimism, pessimism, equally likely decision criterion, criterion of realism, criterion of regret. Decision tree analysis, Decision making with utilities.

Unit–IV

S-S policy for inventory and its derivation in the case of exponential demand; Models with variable supply and models for perishable Items. Replacement Problems; Introduction, block and age replacement policies, replacement of items with long life. Machine interference problems.

Unit-V

Introduction to simulation, generation of random numbers for Uniform, Normal, Exponential, Cauchy and Poisson Distributions. Estimating the reliability of the random numbers, Simulation to Queuing and Inventory problem.

REFERENCES

1. Taha, H.A.(1982): Operations Research : An Introduction; McMillan
2. Kantiswarup;Gupta P.K. and Singh,M.N.(1985) : Operations Research; Sultan Chand.
3. Sharma,S.D.: Operations Research.
4. Sharma J.K : Operation Research
5. Hillier F.S. and Leiberman,G.J.(1962) : Introduction to Operations Research; Holdon Day.
6. Philips,D.T.,Ravindran,A.and Solberg,J.(2000): Operations Research principles and practice.
7. Taha, H.A.(1982): Operations Research : An Introduction; McMillan
8. Kantiswarup;Gupta P.K. and Singh,M.N.(1985) : Operations Research; Sultan Chand.
9. Sharma,S.D.: Operations Research.

M.SC.(STATISTICS) IV-SEMESTER

ELECTIVE-I: STS-303(B): PAPER-III (B): TEXT ANALYTICS (TA)

Unit - I

Introduction to Natural Language Processing Basic, Language Syntax and Structure (Words, Phrases, Clauses, & Grammar), Language Semantics Processing, (Lexical Semantic Relations, Homonyms, Homographs, and Homophones, Capitonyms, Hyponyms and Hypernyms),Text Corpora (Corpora Annotation and Utilities),Accessing Text Corpora (Brown Corpus, WordNet Corpus) and NLP Applications (Machine Translation, Text Summarization and Text categorization).

Unit – II

Concept of the Tokenization, Sentence Tokenization, Word Tokenization, Concept of the Text Normalization, (Cleaning Text, Removing Special characters, Removing stop words,..etc) correcting words using stemming and Lemmatization and Understanding text syntax and structure.(POS tagging and Parsing)

Unit – III

Concepts of feature extraction, Methods of Feature extraction (Bag of words Model, TF-IDF Models, Advanced word Factorization Models likes Word2vec), Strengths and weakness of models and Word cloud … etc, Concepts of Document term matrix, Term Document Matrix.

Unit – IV

Concepts of Topic Modelling, Algorithms of Topic Modelling (Latent Semantic Indexing (LSI) , Latent Dirichlet Allocation (LDA), Non Negative Matrix Factorization (NMF) and Similarity based text clustering models).

Unit-V

Text Classification using supervised methods (Like Multinomial Naïve Bayes, Support vector machines, Random Forest …), concept of Sentiment Analysis and its applications.

REFERENCES

1) Chapman & Hall : Handbook of Natural Language Processing, Second Edition.
2) CRC: Machine Learning & Pattern Recognition, 2nd Edition.
3) Christopher Manning and Hinrich Schuetze: Foundations of Statistical Natural Language Processing.
4) Dipanjan Sarkar: Text Analytics with Python, A press Publication.
5) Julia Silge: Text Mining with R: A Tidy Approach, 1st Edition.

M.SC. (STATISTICS) IV-SEMESTER

ELECTIVE-I: STS-403(C): PAPER III(C): DEMOGRAPHY (DGY)

Unit–I

Coverage and content errors in demographic data, use of balancing equations and Chandrasekharan, Deming formula to check completeness of registration data.

Unit-II

Adjustment of age data - use of Whipple, Myer and UN indices. Population composition, dependency ratio.

Unit–III

Measures of fertility; stochastic models for reproduction, distributions of time to first birth, inter-live birth intervals and of number of births (for both homogeneous and nonhomogeneous groups of women), estimation of parameters; estimation of parity progression ratios from open birth interval data.

Unit–IV

Measures of Mortality; construction of abridged life tables. Distributions of life table functions and their estimation. Stable and quasi-stable populations, intrinsic growth rate. Models for population growth and their fitting to population data. Stochastic models for population growth.

Unit–V

Stochastic models for migration and for social and occupational mobility based on Markov chains. Estimation of measures of mobility. Methods for population projection. Use of Leslie matrix.

REFERENCES

1. Bartholomew, D. J. (1982). Stochastic Models for Social Processes, John Wiley.
2. Benjamin, B. (1969). Demographic Analysis, George, Allen and Unwin.
3. Chiang, C. L. (1968). Introduction to Stochastic Processes in Biostatistics; John Wiley.
4. Cox, P. R. (1970). Demography, Cambridge University Press.
5. Keyfitz, N. (1977). Applied Mathematical Demography; Springer Verlag.

M.SC. (STATISTICS) IV-SEMESTER

ELECTIVE-II:STS-404(A): PAPER IV(A): ARTIFICIAL NEURAL NETWORKS (ANN)

Unit – I

Basics of Artificial Neural Networks (ANN), Human vs Computers, Organization of the Brain, Biological Activations of Neuron; Artificial Neuron Models: McCulloch-Pitts, Perceptron, Adaline, Hebbian Models; Historical Developments of ANN, Characteristics of ANN, Types of Neuron Activation Function, Signal functions and their properties, monotonicity. ANN Architecture, Classification Taxonomy of ANN, Un-supervised and Reinforcement learning; Learning tasks, Memory, Adaptation, Statistical nature of the learning process. Statistical learning theory**.** Gathering and partitioning of data for ANN and its pre and post processing.

Unit – II

Perceptron Learning Algorithm, Derivation, Perceptron convergence theorem (statement); Multi-layer Perceptron Learning rule, limitations. Applications of the Perceptron learning. Gradient Descent Learning, Least Mean Square learning, Widrow-Hoff Learning. Feed-forward and Fedd-back Back-Propagation Algorithms and derivation, learning rate, Momentum, Difficulties and Improvements. Bias and Variance. Under- Fitting and Over-Fitting

Unit-III

Radial Basis Function Networks: Introduction, Regularization theory, Regularization Networks, Generalized Radial Basis Function Networks, Approximation properties of Radial Basis Function Networks, Comparison with Multi-layer Perceptron, Applications.

Unit-IV

Hebbian Learning, Competitive learning**.** Self Organizing Maps**:** Two basic feature mapping models, Self-Organizing Map, SOM algorithm, properties of feature map, computer simulations, Vector quantization, Learning vector quantization, Hierarchical Vector Quantization,

Unit-V

Boltzman Machine and its learning rule, Hopfield model and its learning. Sigmoid belief network learning procedure, Stochastic machines. Applications of ANN in Classification, Clustering, Regression, Time series forecasting.

REFERENCES

1. Haykin, S. (1994). Neural Networks: A Comprehensive Foundation. New York: Macmillan Publishing. A comprehensive book and contains a great deal of background theory
2. Yagnanarayana, B. (1999): “Artificial Neural Networks” PHI
3. Bart Kosko(1997): Neural Networks and Fuzzy systems, PHI
4. Jacek M. *Zurada(1992): Artificial Neural Systems,* West Publishing Company.
5. Carling, A. (1992). Introducing Neural Networks. Wilmslow, UK: Sigma Press.
6. Fausett, L. (1994). Fundamentals of Neural Networks. New York: Prentice Hall.

M.SC. (STATISTICS) IV-SEMESTER

ELECTIVE-II: STS-404(B): PAPER IV(B): DESIGN AND ANALYSIS OF ALGORITHMS

UNIT I

Introduction to Algorithms: Algorithm, Time & space complexity, Asymptotic Notations. Writing pseudocode, Design Techniques.

Divide and Conquer: Control Abstraction, Binary Search, Finding the Maximum and Minimum, Merge Sort; Quick Sort, Selection sort, Strassen's Matrix Multiplication, Convex Hull.

UNIT-II

Greedy Method: Control Abstraction, Knapsack Problem, Job Sequencing with Deadlines, Minimum-Cost Spanning Trees (Kruskal's & Prim's), Single Source Shortest Paths (Dijkstra's).

Dynamic Programming: Control Abstraction, Multistage Graphs, All-Pairs Shortest Paths, Single-Source Shortest Paths, Optimal Binary Search Trees, 0/1 Knapsack, Traveling Salesperson Problem.

UNIT-III

Basic Traversal and Search Techniques: Techniques for Binary Trees, Techniques for Graphs, Connected Components and Spanning Trees, Biconnected Components and DFS.

Back Tracking: Control Abstraction, , 8-Queens Problem, Sum of Subsets, Graph Colouring, Hamiltonian Cycles, Knapsack Problem.

Branch-Bound: Control Abstraction, 0/1 Knapsack Problem, Traveling Sales Person problem.

UNIT -IV

NP-Hard and NP-Complete Problems: Basic Concepts, Cook's Theorem, NP-Hard. Graph Problems, NP-Hard Scheduling Problems, NP-Hard Code Generation, Some Simplified NP-Hard Problems.

REFERENCE BOOKS

1. E Horowitz, S Sahni, S Rajasekaran, (2007): Fundamentals of Computer Algorithms, 2/e, Universities Press.
2. T.H. Cormen, CE Leiserson, R.L Rivert, C Stein, (2010): Introduction to Algorithms, 3/e, PHI.
3. R. Pannerselvam (2007): Design and Analysis of Algorithms, PHI.
4. Hari Mohan Pandey, (2009): Design, Analysis and Algorithm, University Science Press.

M.SC. (STATISTICS) IV-SEMESTER

ELECTIVE-II: STS-404(C): PAPER IV(C): CLINICAL TRIALS (CT)

Unit–I

Introduction to clinical trials : The need and ethics of clinical trials, bias and random error in clinical studies, conduct of clinical trials, overview of Phase I-IV trials, multi-center trials. Data management: data definitions, case report forms, database design, data collection systems for good clinical practice.

Unit-II

Determination of sample size: for two independent samples of Dichotomous Response variables, for two independent samples of Continuous Response variables and for repeated variables.

Unit–III

Design of clinical trials : parallel vs. cross-over designs, cross-sectional vs. longitudinal designs, review of factorial designs, objectives and endpoints of clinical trials, design of Phase I trials, design of single-stage and multi-stage Phase II trials, design and monitoring of Phase III trials with sequential stopping, design of bioequivalence trials.

Unit–IV

Reporting and analysis: analysis of categorical outcomes from Phase I - III trials, analysis of survival data from clinical trials.

Unit–V

Surrogate endpoints: selection and design of trials with surrogate endpoints, analysis of surrogate endpoint data. (2L) Meta-analysis of clinical trials.

REFERENCES

1. S. Piantadosi (1997). Clinical Trials : A Methodological Perspective. Wiley and Sons.
2. C. Jennison and B. W. Turnbull (1999). Group Sequential Methods with Applications to Clinical Trials, CRC Press.
3. L. M. Friedman, C. Furburg, D. L. Demets (1998). Fundamentals of Clinical Trials, Springer Verlag.
4. J. L. Fleiss (1989). The Design and Analysis of Clinical Experiments. Wiley and Sons.
5. E. Marubeni and M. G. Valsecchi (1994). Analyzing Survival Data from Clinical Trials and Observational Studies, Wiley and Sons.

M.SC.(STATISTICS) SEMESTER IV

STS-405: PAPER-V: STOCHASTIC PROCESSES & TIME SERIES ANALYSIS
PRACTICAL-I (CONVENTIONAL)

SECTION-A: LIST OF PRACTICALS ON STOCHASTIC PROCESSES

1. Formulation of problems as Markov chain models
2. Computation of finite dimensional and marginal distributions; higher dimensional transition probabilities.
3. Classification of states, identification of recurrent classes and reduction to canonical form of t.p.m.
4. Probabilities of absorption into recurrent classes (from transient states)
5. Computation of stationary distribution (unique case)
6. Computation of stationary distribution (non-unique case)
7. M|M|1 queue – operating characteristics
8. Mean and variance of n^{th} generation size and probability of extinction of Branching processes.

LIST OF PRACTICALS ON TIME SERIES ANALYSIS

1. Generation of Time series by means of simple time series models
2. Sample and theoretical correlograms
3. Periodogram analysis
4. Writing the models in B notation and stationarity and invertibility of the models
5. Classification of ARIMA models and computation of weights
6. Identification AR, MA, ARMA models
7. Estimation of parameters in AR, MA and ARMA models
8. Computation of forecasts, updating and probability limits for forecasts

M.SC.(STATISTICS) SEMESTER IV

STS-406: PAPER VI: ELECTIVE-I & ELECTIVE-II

PRACTICAL-II (CONVENTIONAL & WITH SOFTWARE)

ELECTIVE-I	ELECTIVE-II
A) Advanced Operations Research (AOR) B) Text Analytics (TA) C) Demography (DGY)	A) Artificial Neural Networks (ANN) B) Design & Analysis of Algorithms (DAA) C) Clinical Trails (CT)

ELECTIVE-I (A): ADVANCED OPERATIONS RESEARCH

1. Wolfe and Beale's methods for QPP
2. Separable Programming problem
3. Dynamic Programming Problem
4. Goal Programming Problem
5. Problems on Decision under uncertainty
6. Replacement Problem

ELECTIVE-I(B): TEXT ANALYTICS (TA)

1. Perform data collection by web scrapping with python and Perform following tasks (i) Find the URL that you want to scrape (ii) Inspecting the Page (iii) Find the data you want to extract (iv) Write the code (v) Run the code and extract the data (vi) Store the data in the required format.
2. Perform following Data Pre-processing tasks in Python using Scikit-learn. standardization, normalization, encoding, discretization, imputation of missing values. Use your own dataset to perform all pre-processing tasks as suggested in given reference.
   (i) https://www.analyticsvidhya.com/blog/2016/07/practical-guide-datapreprocessing-python-scikit-learn/
   (ii) https://scikit-learn.org/stable/modules/preprocessing.html
3. Answer the following question in your blog (As per dataset taken by you): Dataset Description: Task to be performed: How to decide variance threshold in data reduction? Code Snapshot, Output Snapshot, Task-2, Code Snapshot, Output Snapshot.
   Perform following Data Pre-processing tasks using python
   Data reduction using variance threshold, univariate feature selection,
   recursive feature elimination, PCA, correlation
   Reference:
   1. https://medium.com/analytics-vidhya/feature-selection-using-scikit-learn5b4362e0c19b
   2. https://machinelearningmastery.com/rfe-feature-selection-in-python/
   3. https://towardsdatascience.com/pca-using-python-scikit-learn-e653f8989e60
   4. https://towardsdatascience.com/feature-selection-using-python-for-classificationproblem-b5f00a1c7028
   5. https://www.analyticsvidhya.com/blog/2016/01/guide-data-exploration/
   Answer the following question in your blog (As per dataset taken by you):
   Dataset Description; Task to be performed; Why feature selection is important?? Its advantages/disadvantages. Code Snapshot; Output Snapshot;What is the impact on accuracy, with or without data reduction? Code Snapshot; Output Snapshot.
   Amongst all methods, which method avoids overfitting and improves model performance

ELECTIVE– I(C): DEMOGRAPHY

1. Construction of Abridged life tables
2. Fitting of population growth models
3. Estimation of population projection
4. Estimation of Life table functions

Elective-II (A): Artificial Neural Networks (ANN)

1. Implementation of Perceptron Learning Algorithm.
2. Implementation of Multi-layer Perceptron Learning
3. Implementation of Gradient Descent Learning,
4. Implementation of Least Mean Square learning,
5. Implementation of Widrow-Hoff Learning.
6. Implementation of Back-Propagation Algorithms.
7. Implementation of Markov Decision Process,
8. Implementation of Hidden Markov Model,

Elective-II(B): DESIGN & ANALYSIS OF ALGORITHMS (Using Python)

1. Write a program for sorting the given list using: Merge Sort, Quick Sort, Heap Sort.
2. Write a program to find the given number in a list using Binary Search.
3. Write a program to find the minimal spanning tree using Kruskal's and Prims Algorithms.
4. Write a program to find the shortest path using Dijkstra's Algorithm.
5. Write a program to solve using dynamic programming technique for Travelling sales man problem. Multistage Graph problem, Optimal Binary Search Trees.
6. Write a program to solve Knapsack problem using Back tracking

ELECTVE-II (C): CLINICAL TRIALS

1. Determination of Sample size
2. Multiple Logistic Regression with two or Three variables
3. Analysis of Clinical trial data using Cross over design
4. Analysis of Clinical trial data using Parallel design
5. Meta-analysis of Clinical trials
6. Analysis of Clinical data using Factorial Experiments

M.SC. (STATISTICS) SEMESTER-IV

STS-407: PAPER VII: MAJOR PROJECT

Note: Follow the guidelines of the project specified in STS-308: Mini project.

****