"Dissemination of Education for Knowledge, Science and Culture"
-Shikshanmaharshi Dr. Bapuji Salunkhe

Shri Swami Vivekanand Shikshan Sanstha's

Vivekanand College Kolhapur (Empowered Autonomous)

Department of Statistics

M. Sc. II- Applied Statistics

Semester III & IV

Syllabus to be implemented from
Academic year 2024-25
(As per NEP 2020)

Teaching and Evaluation scheme

One/Two- Years PG Programme

Department/Subject Specific Core or Major (DSC)

Second Year Semester- III & IV (M. Sc. Applied Statistics)

Semester	Course Code	Course Title	No. of Credits
III	DSC18STA31	Stochastic Processes	**04**
	DSC18STA32	Statistical Learning and Data Mining	**04**
	DSC18STA33	Python Programming	**02**
	DSE18STA31	Biostatistics	**04**
	DSE18STA32	Statistical Quality Control	**04**
	RPR18STA31	Research Project	**04**
	DSC18STA39	Statistics Practical III	**04**
		Credit	**22**
IV	DSC18STA41	Time Series Analysis	**04**
	DSC18STA42	Design and Analysis of Experiments	**04**
	DSE18STA41	Optimization Techniques	**04**
	DSE18STA42	Python for Data Science	**04**
	RPR18STA41	Research Project	**06**
	DSC18STA49	Statistics Practical IV	**04**
		Credits	**22**

1)

Nature of the theory question papers (4 credits):

a) There shall be 7 questions each carrying 16 marks.

b) Question No.1 is compulsory.

i) It consists of 8 multiple choice questions for 1 mark each. (8 marks)

ii) short note type questions (attempt any 2 out of 3) (8 marks)

c) Students have to attempt any 4 questions from question No. 2 to 7.

d) Question No. 2 to 7 shall contain 2 to 4 sub-questions.

2) Nature of the theory question papers (2 credits):

a) There shall be 4 questions.

b) Question No.1 is compulsory.

i) It consists of 4 multiple choice questions for 1 mark each. (4 marks)

ii) short note type questions (attempt any 1 out of 2) (4 marks)

c) Question No. 2 to 4 shall be of 16 marks each.

d) Students have to attempt any 2 questions from question No. 2 to 4.

e) Question No. 2 to 4 shall contain 2 to 4 sub-questions.

3)Nature of Practical examination: -

Sr. No.	Component	Max marks
1	Practical examination: Examination will be of 3-hour duration. There shall be 8 questions each of 12 marks, of which a student has to attempt any 5 questions.	60
2	Day-to-day practical performance and journal	20
3	Viva: Viva will be based on all practical's	20
Total		100

4)Each of the following courses have the same question paper in all examinations of M.Sc. Statistics and M.Sc. Applied Statistics.

Semester III	Semester IV
Stochastic Processes	Time Series Analysis
Statistical learning and Data Mining	Design and Analysis of Experiments
Python Programming	-

DSC18STA31: STOCHASTIC PROCESSES

Course Objectives: At the end of the course students will be able to:

CO1: Classify general Stochastic Process.

CO2: Learn random walk model, Gambler Ruin Problem and able to compute long run distribution of Markov chain.

CO3: Apply the Poisson process, Birth and Death process and various Queuing systems in real life.

CO4: Employ Branching processes to solve real life problems of stochastic nature

Unit and Credit	DSC18STA31: STOCHASTIC PROCESSES	No. of hours per unit / credits
Unit I	Definition of stochastic process, classification of stochastic processes according to state space and time domain, finite dimensional distributions. Examples of various stochastic processes. Definition of Markov chain. Examples of Markov chains, Formulation of Markov chain models, initial distribution, transition probability matrix, Chapman-Kolmogorov equations, calculation of n-step transition probabilities. Simulation of Markov Chain.	15
Unit II	Classification of states, irreducible Markov chain, period of the state, random walk and gambler's ruin problem, first entrance theorem, first passage time distribution. Long-Run proportions and limiting probabilities, relation with mean recurrence time, stationary distribution.	15
Unit III	Discrete state space continuous time Markov chain, Poisson process and related results. Birth and death processes and associated cases. Renewal and delayed renewal processes, related theorems, key renewal theorem (Without proof) and its application. Simulation of Poisson process and discrete state space Markov processes.	15
Unit IV	Galton-Watson Bienayame Branching process. Generating functions and its properties, moments. Probability of ultimate extinction. Distribution of population size and association results. Simulation of branching process, Ehrenfest model. Basic elements of Queuing model. Steady state probabilities and various average characteristics for the models: M/M/1, M/M/1 with balking, M/M/c and M/G/1.	15

Books Recommended:

1. Bhat B. R. (2000). Stochastic Models: Analysis and Applications, (New Age International)
2. Cinlar E. (2013): Introduction to Stochastic Process. (Courier Corporation)
3. Feller W.(2008): An Introduction to Probability Theory and Its Applications. (Wiley)
4. Hoel P. G., Port S. C. and Stone C. J. (1987): Introduction to Stochastic Processes. (Waveland Press)
5. Karlin S. and Taylor H. M. (1968): A First Course in Stochastic Process. (Academic Press)
6. Medhi J. (2009): Stochastic Process, (New Age International Publications)
7. Ross S. (1996): Stochastic Processes. (Wiley)
8. Ross S. (2014): Introduction to Probability Models. (Academic Press)
9. Taylor H. M. and Karlin S. (2014): An Introduction to Stochastic Modeling (Academic Press)

DSC18STA32: Statistical Learning and Data Mining

Course Objectives: At the end of the course students will be able to:
CO1: Understand and clean the big data sets.

CO2: Apply classification methods to real life problems in various fields.

CO3: Select and evaluate the models based on datasets using different modelling techniques.
CO4: Understand unsupervised learning and supervised learning techniques for univariate and multivariate data.

Unit and Credit	DSC18STA32: Statistical Learning and Data Mining	No. of hoursper unit / credits
Unit I	Data understanding and data cleaning, concept of supervised and unsupervised learning. Problem of classification, classification techniques: k-nearest neighbor, decision tree, Naïve Bayesian, classification based on logistic regression, Bayesian belief Network.	15
Unit II	Model evaluation and selection: Metrics for Evaluating Classifier Performance, Holdout Method and Random Subsampling, Cross-Validation, Bootstrap, Model Selection Using Statistical Tests of Significance, Comparing Classifiers Based on Cost–Benefit and ROC Curves. Techniques to Improve Classification Accuracy: Introduction to Ensemble Methods, Bagging, Boosting and AdaBoost, Random Forests, Improving Classification Accuracy of Class-Imbalanced Data.	15
Unit III	ANN and SVM: Artificial Neutral Network (ANN): Introduction to ANN, types of activation function, McCulloch-Pitts AN model, single layer network, multilayer feed forward network model, training methods, ANN & regression models. Support vector machine: Introduction to support vector machine, loss functions, soft margin, optimization hyperplane, support vector classification, support vector regression, linear programming support vector machine for classification and regression.	15
Unit IV	Unsupervised learning: Clustering: k-mediods, CLARA, DENCLUE, DBSCAN, Probabilistic model based clustering, Scalable clustering, Market Basket Analysis: Association rules and prediction, Apriori Algorithm, data attributes, applications to electronic commerce.	15

Books Recommended:

1. Berson and Smith S.J. (1997) : Data warehousing, Data Mining, and OLAP, McGraw-Hill.
2. Breiman J.H Friedman, R.A. Olshen and stone C.J. (1984) : Classification and Regression Trees, Wadsworth and Brooks / Cole.
3. Han, J. and Kamber, M. and Pei, J. (2012) : Data Mining: Concepts and Techniques. MorganGaufmann.3rd Edition.
4. Mitchell T.M. (1997) : Machine Learning , McGraw-Hill.
5. Ripley B.D. (1996) : Pattern Recognition and Neural Networks. Cambridge University Press.

6. Vapnik V.N. The nature of Statistical learning theory, Springer.
7. Cristianini N. and Shawe-Taylor J. An Introduction to support vectormachines.
8. Data set source: http://www.ICS.uci.edu/~mlearn/MLRepository.html
9. Mehrika, K., Mohan, C., and Ranka (1997) Elements of Artificial neural networks. Penram international.
10. Hastie T, Tibshirani R, Friedmant J, (2009): The elements of statistical Learning, Springer.
11. Chattamvelli, R. (2015). Data mining methods. Alpha Science International.

DSC18STA33: Python Programming

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

CO1: Understand Python programming basics, utilizing Python libraries.
CO2: Write a programme in python.
CO3: Understand use of functions and file handling in Python.
CO4: Analyze data using Python.

Unit	DSC18STA33: Python Programming	Hours Allotted
1	Introduction to Python, History of Python, Introduction to Python Interpreter and program execution, Python Installation Process, Introduction to anaconda, python variable declaration, Keywords, Indents in Python, Python input/output operations. Types of Operators, Built-in Data types: Arrays, String, List, Tuple, Set, Dictionary (characteristics and methods). Conditional Statements & Loop Conditional Statements (If, If-else, If-elif-else, Nested-if etc.) and loop control statements (for, while, Nested loops, Break, Continue, Pass statements). List comprehensions. Function in python: Introduction to functions, Function definition and calling, Function parameters, Default argument function, Variable argument function, in built functions in python, lambda function, local and global variables. File Processing Concept of Files, File opening in various modes and closing of a file, Reading from a file, Writing onto a file, some important file handling functions e.g open(), close(), read(), readline() etc.	10
2	Modules: Concept of modularization, Importance of modules in python, importing modules, Built in modules. Concept of library and its working, Data storage, manipulation, visualization and analysis using the libraries: Numpy, Pandas, Scipy, statsmodels, Matplotlip, Seaborn, Regular Expressions (RegEx), Ski-kit learn.	20

Reference Books:

1. Practical Programming: An introduction to Computer Science Using Python, second edition, Paul Gries, Jennifer Campbell, Jason Montojo, The Pragmatic Bookshelf.
2. Python for Informatics: Exploring Information, Charles Severance
3. Learning Python, Fourth Edition, Mark Lutz, O'Reilly publication
4. Introduction to Python for Computational Science and Engineering (A beginner's guide), Hans Fangohr
5. John V Guttag. "Introduction to Computation and Programming Using Python", Prentice Hall of India
6. R. Nageswara Rao, "Core Python Programming", Dreamtech

DSE18STA31: Biostatistics

Course Objectives: At the end of the course students will be able to:

CO1: Understand need and ethics of any clinical trial and how to conduct clinical trial of any medicine in different phases.

CO2: Apply various designs of clinical trials to the data.

CO3: Understand the designs of clinical trials and Epidemiological studies.

CO4: Analyze and report the clinical trials.

Unit and Credit	DSE18STA31: Biostatistics	No. of hours per unit / credits
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-centre trials. Data management: data definitions, case report forms, database design, data collection systems for good clinical practice, concept of blinding/masking in clinical trials. Bioavailability, pharmacokinetics and pharmaco-dynamics, two compartment model.	15
Unit II	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.	15
Unit III	Design of bio-equivalence trials, Inference for 2x2 crossover design: Classical methods of interval hypothesis testing for bioequivalence, Bayesian methods, nonparametric methods. Reporting and analysis: analysis of categorical outcomes from Phase I - III trials, analysis of survival data from clinical trials.	15
Unit IV	Epidemiological studies: case-control and cohort designs. Measures of disease occurrence and association, variation and bias, identifying non-causal association and confounding, communicating results of epidemiological studies, ethical issues in epidemiology. Causal Inference.	15

Books Recommended:

1. C. Jennison and B. W. Turnbull (1999): Group Sequential Methods with Applications to Clinical Trials, CRC Press.
2. Chow S.C. and Liu J.P. (2004). Design and Analysis of Clinical Trials. 2nd Ed. Marcel Dekkar.
3. Chow S.C. and Liu J.P.(2009). Design and Analysis of Bioavailability and bioequivalence. 3rd Ed. CRC Press.
4. Clayton, D. and Hills, M. (2013). Statistical methods in Epidemiology, OUP.
5. Daniel, W. W. and Cross, C. L. (2012). Biostatistics: A Foundation for Analysis in the Health Sciences, 10th Edition, Wiley.
6. J. L. Fleiss (1989). The Design and Analysis of Clinical Experiments. Wiley and Sons.

DSE18STA32: Statistical Quality Control

Course Objectives: At the end of the course students will be able to:
CO1: Understand concepts of control charts in quality improvement.
CO2: Construct modified control charts to monitor the process.
CO3: Analyse process capability using control charts.
CO4: Evaluate the quality of products using various acceptance sampling plans.

Unit and Credit	DSE18STA32: Statistical Quality Control	No. of hours per unit / credits
Unit I	Quality Improvement Tools: affinity diagram, interrelationship digraph, tree diagram, prioritization matrix, matrix diagram, process decision program chart, activity network diagram, stem-and-leaf display, dot diagrams, boxplot, and normal probability plot, Engineering Process Control. Shewhart Control charts: basic statistical principles and assumptions, phase I and phase II applications, benefits from the use of control charts, concept of rational subgroups, performance measures of a control chart, X, R, S, S2, p, c and D charts, σ-control limits and probability control limits, over dispersion. Modifications to control chart procedures: warning limits, sensitizing rules, adaptive design parameters, integration of two charts. Concept of economic design of a control chart.	15
Unit II	Alternatives to Shewhart control charts: CUSUM and EWMA charts, Shewhart–EWMA Chart. Multivariate Control Charts: multivariate chart versus individual charts, Hotelling's T2 control chart, multivariate CUSUM Charts, multivariate EWMA charts, Regression adjustment. Other Control Charts: SPRT chart, GLR Chart, charts for autocorrelated data, nonparametric control charts, Bayesian control charts. The change point model for process monitoring.	15
Unit III	Process capability Analysis: process capability, process capability indices (Cp, Cpk, Cpm, Cpmk), point and interval estimation of Cp and Cpk, Nonparametric Capability Indices: robust capability indices, capability indices based on fitted distributions, data transformation, capability indices computed using resampling methods. Multivariate Process Capability Indices. Six Sigma Methodology: components of a Six Sigma methodology, the DIMAC process, Six Sigma applications, Six Sigma concept for customer satisfaction, Six Sigma training, Lean Six Sigma.	15

Unit IV	Acceptance sampling plans for attributes: single sampling plan, double and multiple sampling plans, sequential sampling. Performance measures of sampling plans. Acceptance sampling plans for variables: Advantages and Disadvantages of Variables Sampling, Sampling inspection plans by variables for one or two sided specifications, Sequential Sampling by Variables, Rectifying inspection of lots, the Deming inspection criterion, Continuous sampling plans, skip-lot sampling plans.	15

Books Recommended:

1. Guenther, W. C. (1977). Sampling Inspection in statistical quality control. Macmillan.
2. Kenett, R. S. and Zacks, S. (2014). Modern Industrial Statistics with applications in R, MINITAB and JMP. John Wiley & Sons.
3. Montgomery, D. C. (2010). Statistical Quality Control: A Modern Introduction, 6th Edition. Wiley India Pvt Ltd.
4. Ryan, T. P. (2011). Statistical Methods for Quality Improvement. John Wiley & Sons.

DSC18STA39: STATISTICS PRACTICAL -III

Practical Number	Practical Name
1	Simulation of Markov chain
2	Classification of t.p.m. and computation of n- step probability matrix.
3	Classification of states: Computations of absorption probabilities.
4	Stationary distribution and recurrence time.
5	Classification.
6	Cluster Analysis.
7	Artificial Neural Network.
8	Support Vector Machine.
9	Python Programming I (Arrays, String, List, Tuple, Set, Dictionary)
10	Python Programming II (Conditional statements, Loops, List comprehensions)
11	Python Programming III (File processing, Modules)
12	Python Programming IV (Data visualization, manipulation using python libraries)
13-16	At least 3-4 practicals on elective paper.

(Each practical should consist of problems to be solved using at least two of the following software: EXCEL/ R/python)

Sem-IV

DSC18STA41: Time Series Analysis

Course Objectives: At the end of the course students will be able to:

CO1: Remove trend and seasonality using different methods to convert the time series into stationary.

CO2: Understand time series, auto-covariance, auto-correlation their properties, various smoothing techniques.

CO3: Obtain Causality and inevitability, π -weights and ψ- weights, ACVF, ACF, PACF.

CO4: Understand estimation of ARIMA model, residual analysis and diagnostic checking, their forecasting.

Unit and Credit	DSC18STA41: Time Series Analysis	No. of hoursper unit / credits
Unit I	(a) Exploratory time series analysis, Exponential, Double exponential and Holt – Winter smoothing and forecasting. (b) Auto - Covariance, Auto-correlation functions and their properties and characterization (without proof), Partial auto covariance function, Auto – Covariance generating function. First and second order Stationary time series, white noise process, Linear Process, Sample Estimates of mean, Auto - Covariance, Auto-correlation and Partial auto covariance functions.	15
Unit II	Wold representation of linear stationary processes, linear time series models: Autoregressive, Moving Average, Autoregressive Moving Average models. Causality and invertibility of ARMA processes, computation of π -weights and ψ- weights, computation of ACVF, ACF and PACF for AR(1), AR(2), MA(1), MA(2), ARMA(1,1) process and general procedure for ARMA(p,q) process. The need for differencing a time series, Autoregressive Integrated Moving Average models.	15
Unit III	(a) Estimation of ARMA models: Yule-Walker estimation for AR Processes, Maximum likelihood and least squares estimation for ARMA Processes, Residual analysis and diagnostic checking. Minimum mean squared error forecasting for ARMA and ARIMA models, updating forecasts. Introduction to SARIMA models. (b) Spectral Representation of the ACVF, Spectral density of an ARMA process, its computation for simple models.	15
Unit IV	(a) Introduction to ARCH and GARCH models. Properties and estimation under ARCH(1) and GARCH(1,1) model. (b) Vector time-series models: Covariance and Correlation Matrix functions, MA and AR representation of vector processes, Covariance matrix function of the vector AR(1) and MA(1) models.	15

Books Recommended:

1. W. S. Wei (2005) Time Series Analysis: Univariate and Multivariate Methods
2. Box, G.E.P and Jenkins G.M. (1970) Time Series Analysis, Forecasting & Control, Holden-Day.
3. Brockwell, P.J and Davis R.A. (1987) Time Series: Theory and Methods, Springer-
4. TsayR. S. Analysis of Financial Time Series, 3rd Ed. (Wil. Ser. in Prob. and Statistics)
5. Kendall, M.G. (1978) Time Series, CharlerGraffin
6. Chatfield, C. (2004) The Analysis of Time Series - An Introduction, Sixth edition, Chapman and Hall.

DSC17STA42: DESIGN AND ANALYSIS OF EXPERIMENTS

Course Objectives: At the end of the course students will be able to:

CO1: Recognize the basic concepts of design of experiments.

CO2: Analyze different factorial and fractional experiments their interactions, graphical representation and confounding.

CO3: Describe the concept of fractional factorial design.

CO4: Understand the concept of response surface and methods of fitting it.

Unit and Credit	**DSC17STA42: DESIGN AND ANALYSIS OF EXPERIMENTS**	**No. of hours per unit / credits**
Unit I	Concept of design of experiments (DOE), applications of DOE; Basic principles of DOE; one-way ANOVA, two-way ANOVA with and without interaction, two-way ANOVA with r observations per cell, Contrasts, orthogonal contrasts, Scheffe's method for comparing contrasts; Comparing pairs of treatment means: controlling false discovery rate, Tukey's test, Fisher least significant difference method; Comparing treatment means with a control; General block design, balanced incomplete block design using fixed effect models and estimation of the model parameters.	
Unit II	Concepts of factorial designs, main effects, and interaction effects; The two-factor factorial design and its analysis using fixed effect model; The general factorial design; Analysis of replicated and unreplicated 2^k full factorial designs; Blocking and confounding in a 2^k factorial design; Construction and analysis of 2^{k-p} fractional factorial designs and their alias structures; Design resolution, resolution III and resolution IV designs; fold over designs; saturated designs.	15
Unit III	The 3^k full factorial design and its analysis using fixed effect model; Confounding in 3^k factorial designs; Construction and analysis of 3^{k-p} fractional factorial designs and their alias structures; Factorials with mixed levels: factors at two and three levels, factors at two and four levels; Design optimality criteria	15

Unit IV	Response surface methodology: the method of steepest ascent, analysis of the response surface using first and second order model, characterizing the response surface, ridge systems, multiple responses, designs for fitting response surfaces: simplex design, central composite design (CCD), spherical CCD, Box–Behnken design; Robust parameter design: crossed array designs and their analyses, combined array designs and the response model approach; The concepts of nested and split-plot designs.	15

Books Recommended:

1.Montgomery D.C. (2017): *Design and Analysis of Experiments*, 9th edition, John Wiley & Sons, Inc.

2. Phadke, M. S. (1989). *Quality Engineering using Robust Design*, Prentice-Hall.
3. Voss, D., Dean, A., and Dean, A. (1999). *Design and Analysis of Experiments*, Springer verlag Gmbh.
4. Wu, C. F., Hamada M. S. (2000). *Experiments: Planning, Analysis and Parameter Design Optimization, 2nd edition*, John Wiley & Sons.

DSE18STA41: Optimization Techniques

Course Objectives: At the end of the course students will be able to:

CO1: Understand basics and formulation of linear programming problems and appreciate their limitations; solve linear programming problems using graphical method.

CO2: Apply simplex method to solve real life problems.

CO3: Solve artificial variable technique, duality theory, revised simplex method, sensitivity analysis.

CO4: Understand the concept of Game theory and dynamic programming to solve their problems and understand their real-life applications

Unit and Credit	**DSE18STA41: Optimization Techniques**	**No. of hoursper unit / credits**
Unit I	Convex Sets and Functions: Convex sets, supporting and separating hyperplanes, convex polyhedra and polytope, extreme points, convex functions. Linear programming problem (LPP): Definition and applications, theorems related to the development of simplex algorithm, theorems related to a basic feasible solution, reduction of a feasible solution to a basic feasible solution, improvement of a basic feasible solution, existence of unbounded solution, optimality conditions and other related theorems (statements only), Examples. Artificial variable technique: Two phase method, Big M method, degeneracy.	15
Unit II	Concept of Duality, related theorems, complementary slackness property and development of dual simplex algorithm. Sensitivity Analysis: Changes in the cost vector, requirement vector and non-basic activity vector; addition of new variables and addition of new constraints.	15
Unit III	Integer Linear Programming Problem (ILPP): The concept of cutting plane, cutting plane method for all ILPP and mixed ILLP, Branch and Bound method. Quadratic programming: KuhnTucker conditions, methods due to Beale,	15
Unit IV	Theory of games: two-person zero sum games, minimax and maximin principles, Saddle point, mixed strategies; rules of dominance, solution of 2 x 2 game by algebraic method, Graphical method, Reduction of the game problem as LPP. Dynamic Programming: The Recursion Equation Approach, Computational Procedure, Characteristics of Dynamic Programming, Solution of L.P.P. by Dynamic Programming.	15

Books Recommended:

1. Hadley G.(1969): Linear Programming, Addison Wesley
2. Taha H. A. (1971): Operation Research: An Introduction, Macmillan N.Y.
3. KantiSwaroop& Gupta M. M.(1985): Operations Research, Sultan Chand & Co. ltd.
4. P.Gupta&D.S.Hira(2010): Operation Research, Sultan Chand & Co. ltd.
5. J. K. Sharma. (2003): Operation Research: Theory and Applications. Macmillan.

DSE18STA42: Python for Data Science

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

CO1: Develop algorithmic solutions to data science related problems.

CO2: Analyze the layers in the architecture of convolution neural networks.

CO3: Understand and master basic knowledge, theories and methods in computer vision.

CO4: Comprehend the concepts of natural language processing, its applications and language modelling techniques.

Unit	DSE18STA42: Python for Data Science	Hours Allotted
1	Introduction to Data Science, Different Sectors in Data Science, Basic terminologies in Data Science. Overview of Data storage, manipulation, visualization and analysis using the libraries: Numpy, Pandas, Scipy, statsmodels, Matplotlip, Seaborn, Regular Expressions (RegEx), Ski-kit learn. Overview of neural network concept, optimizers (Stochastic gradient descent, Adaptive Gradient Algorithm (AdaGrad)).	15
2	Machine learning using scikit-learn library: Classification, Regression, and Clustering. Introduction to deep learning, understanding different types of layers in sequential method: Dense, Convolutional Layers, Recurrent Layers, Normalization Layers. Deep Neural Network: architecture design, optimization for Deep NN, regularization methods for deep NN, Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short Term Memory (LSTM), Deep learning Python libraries: tensorflow (Low level) and Keras (High Level).	15
3	Computer vision: Introduction, OpenCv library in Python, Getting Started with images, Basic Operations on Images, Arithmetic Operations on Images, Image Preprocessing: changing colorspaces, geometric transformations, thresholding, smoothing, morphological transformations, gradients, Canny Edge detection, image pyramids, image segmentation with Watershed algorithm, Feature Detection and Description. Image Detection and recognition examples.	15
4	Introduction to Natural Language Processing (NLP), Natural Language Toolkit (NLTK) in Python. Data Preparation: punctuation removal, stop-words removal, numeric value removal, frequent words removal, rare words removal, spelling correction, tokenization, stemming, lemmatization. Feature Engineering: count vectors as features, Term Frequency-Inverse Document Frequency (TF-IDF), TF-IDF vectors as features, word level TF-IDF, N-Gram level TF-IDF, Character level TF-IDF, world cloud, Inverse Document Frequency, word embedding as features, Text/NLP based features, Topic Models as features, word2vec, sentiment analysis.	15

Reference Books:

1. Robert Johansson, Numerical Python – Scientific Computing and Data Science Applications with NumPy, SciPy and Matplotlib, Apress, 2019.
2. Robert Sedgewick, Kevin Wayne, Robert Dondero, Introduction to Programming in Python: An Inter-disciplinary Approach, Pearson India Education Services Pvt. Ltd., 2016

3. Nelli, F., Python Data Analytics: with Pandas, NumPy and Matplotlib, Apress, 2018.
4. Practical Programming: An introduction to Computer Science Using Python, second edition, Paul Gries, Jennifer Campbell, Jason Montojo, The Pragmatic Bookshelf.
5. Python for Informatics: Exploring Information, Charles Severance

DSC18STA49: STATISTICS PRACTICAL -IV

Practical Number	Practical Name
1	Auto covariance and Autocorrelation.
2	Causal and Invertible
3	Smoothing the series
4	Forecasting
5	Analysis of CRD, RBD, LSD
6	Analysis of BIBD
7	Analysis of full, confounded and fractional 2^k factorial designs
8	Analysis of full, confounded and fractional 3k factorial designs
9-13	At least 3-4 practicals on elective paper.

(Each practical should consist of problems to be solved using at least two of the following software: EXCEL/ R/python)

VIVEKANAND COLLEGE KOLHAPUR (Empowered Autonomous) ESTD JUNE 1964

HEAD
DEPARTMENT OF STATISTICS
VIVEKANAND COLLEGE, KOLHAPUR
(EMPOWERED AUTONOMOUS)