# M.Sc Applied Statistics and Data Analytics

## Curriculum

*(effective from the academic year 2020-21)*

### Semester I

<table>
<thead>
<tr>
<th>Course code</th>
<th>Course</th>
<th>L</th>
<th>T</th>
<th>P</th>
<th>Credit</th>
<th>ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>20MAT505</td>
<td>Linear Algebra</td>
<td>3</td>
<td>0</td>
<td></td>
<td>4</td>
<td>A</td>
</tr>
<tr>
<td>20MAT506</td>
<td>Probability Theory and Estimation</td>
<td>3</td>
<td>0</td>
<td></td>
<td>4</td>
<td>B</td>
</tr>
<tr>
<td>20MAT507</td>
<td>Data Structures and Algorithms</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>C</td>
</tr>
<tr>
<td>20MAT508</td>
<td>Optimization Techniques</td>
<td>3</td>
<td>0</td>
<td></td>
<td>4</td>
<td>D</td>
</tr>
<tr>
<td>20MAT509</td>
<td>Introduction to Data Science</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>E</td>
</tr>
<tr>
<td>20MAT510</td>
<td>Python Programming</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>F</td>
</tr>
<tr>
<td>18CUL501</td>
<td>Cultural Education</td>
<td>2</td>
<td>0</td>
<td></td>
<td>P/F</td>
<td>G</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>

### Semester II

<table>
<thead>
<tr>
<th>Course code</th>
<th>Course</th>
<th>L</th>
<th>T</th>
<th>P</th>
<th>Credit</th>
<th>ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>20MAT515</td>
<td>Statistical Inference and Design of Experiments</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>A</td>
</tr>
<tr>
<td>20MAT516</td>
<td>Multivariate Statistics and Regression Analysis</td>
<td>3</td>
<td>1</td>
<td></td>
<td>4</td>
<td>B</td>
</tr>
<tr>
<td>20MAT517</td>
<td>Machine Learning</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>C</td>
</tr>
<tr>
<td>20MAT518</td>
<td>Big Data Analytics and Hadoop</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>D</td>
</tr>
<tr>
<td>20MAT519</td>
<td>Data Mining</td>
<td>3</td>
<td>1</td>
<td></td>
<td>4</td>
<td>E</td>
</tr>
<tr>
<td>20MAT520</td>
<td>Data Security</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>F</td>
</tr>
<tr>
<td>18AVP501</td>
<td>Amrita Value Programme</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>G</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>

### Semester III

<table>
<thead>
<tr>
<th>Course code</th>
<th>Course</th>
<th>L</th>
<th>T</th>
<th>P</th>
<th>Credit</th>
<th>ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>20MAT606</td>
<td>Statistical Quality Control and Reliability</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>A</td>
</tr>
<tr>
<td>20MAT607</td>
<td>Introduction to Deep Learning</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>Elective I</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>Elective II</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>E</td>
</tr>
<tr>
<td></td>
<td>Elective II</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>F</td>
</tr>
<tr>
<td>20MAT690</td>
<td>Live-in-Lab/@/ Open Elective*</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>J</td>
</tr>
<tr>
<td>Course code</td>
<td>Course</td>
<td>L</td>
<td>T</td>
<td>P</td>
<td>Credit</td>
<td>ES</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>--------</td>
<td>----</td>
</tr>
<tr>
<td>20MAT696</td>
<td>Dissertation</td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>P</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

**Total credits for the programme: 77**

**ELECTIVES (any three)**

<table>
<thead>
<tr>
<th>Course code</th>
<th>Course</th>
<th>L</th>
<th>T</th>
<th>P</th>
<th>Credit</th>
<th>ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>20MAT651</td>
<td>Taugchi Techniques</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>D/E</td>
</tr>
<tr>
<td>20MAT652</td>
<td>Special Distribution Functions</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>D/E</td>
</tr>
<tr>
<td>20MAT653</td>
<td>Pattern Recognition</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>D/E</td>
</tr>
<tr>
<td>20MAT654</td>
<td>Stochastic Process</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>D/E</td>
</tr>
<tr>
<td>20MAT655</td>
<td>Queuing Theory</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>D/E</td>
</tr>
<tr>
<td>20MAT656</td>
<td>Market Analytics</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>D/E</td>
</tr>
<tr>
<td>20MAT657</td>
<td>Survival Analysis</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>D/E</td>
</tr>
<tr>
<td>20MAT658</td>
<td>Sampling Techniques</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>D/E</td>
</tr>
<tr>
<td>20MAT659</td>
<td>Demography and Actuarial Statistics</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>D/E</td>
</tr>
<tr>
<td>20MAT660</td>
<td>Official Statistics</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>D/E</td>
</tr>
<tr>
<td>20MAT661</td>
<td>Healthcare Analytics</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>D/E</td>
</tr>
<tr>
<td>20MAT662</td>
<td>Computational Biology</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>D/E</td>
</tr>
<tr>
<td>20MAT663</td>
<td>Computer aided drug designing</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>D/E</td>
</tr>
<tr>
<td>20MAT664</td>
<td>Reinforcement Learning</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>D/E</td>
</tr>
<tr>
<td>20MAT665</td>
<td>Social Network Analytics</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>D/E</td>
</tr>
<tr>
<td>20MAT666</td>
<td>Mining of Massive Datasets</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>D/E</td>
</tr>
<tr>
<td>20MAT667</td>
<td>Parallel and Distributed Systems</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>D/E</td>
</tr>
</tbody>
</table>

*One Open Elective course is to be taken by each student, in the third semester, from the list of Open electives offered by the School.

*Students undertaking and registering for a Live-in-Lab project, can be exempted from registering for the Open Elective course in the third semester.*
M.Sc Applied Statistics and Data Analytics
Syllabus
2020 Admissions onwards

SEMESTER-I

20MAT505 Linear Algebra 310 4

Unit-I

Unit-II

Unit-III
Linear Transformations: Positive definite matrices - Matrix norm and condition number - QR-Decomposition - Linear transformation - Relation between matrices and linear transformations - Kernel and range of a linear transformation - Change of basis - Nilpotent transformations - Trace and Transpose, Determinants, Symmetric and Skew Symmetric Matrices, Adjoint and Hermitian Adjoint of a Matrix, Hermitian, Unitary and Normal Transformations, Self Adjoint and Normal Transformations, Real Quadratic Forms.

Unit-IV
Eigen values and Eigen vectors: Problems in Eigen Values and Eigen Vectors, Diagonalization, Orthogonal Diagonalization, Quadratic Forms, Diagonalizing Quadratic Forms, Conic Sections, Similarity of linear transformations - Diagonalisation and its applications - Jordan form and rational canonical form.

Unit-V
Decompositions : LU,QR and SVD

Text Books


Reference Books:


20MAT506 Probability Theory and Estimation 31 0 4

Course outcomes

CO1: Understand the basics of probability, random variables and distribution functions.
CO2: Gain knowledge about standard statistical distributions and their properties
CO3: Know the importance of two dimensional random variables and correlation studies
CO4: To gain knowledge point estimation and properties
CO5: To gain knowledge about sampling distributions interval estimations.

Unit-I
Review of probability concepts - conditional probability- Bayes theorem.
Random Variable and Distributions: Introduction to random variable – discrete and continuous random variables and its distribution functions- mathematical expectations – moment generating function and characteristic function -

Unit-II
Standard distributions - Binomial, Multinomial, Poisson, Uniform, exponential, Weibull, Gamma, Beta, Normal. Mean, variance and applications of these distributions- Chebyshev’s theorem and central limit theorem.

Unit-III
Joint, marginal and conditional probability distributions for discrete and continuous cases, stochastic independence, expectation of two dimensional random variables, conditional mean and variance, correlation and introduction to regression.

Unit-IV
Point estimation, properties, methods of estimating a point estimator, minimum risk estimators Sampling distributions of mean and variance, Central and Non-central distributions of t, F and Chi-Square distribution. Central limit theorem.

Unit-V
Interval estimation- Confidence interval for one mean, difference of two means, single proportion, difference of two proportions, single variance, ratio of two variances.

TEXT BOOKS /REFERENCE BOOKS:

20MAT507  Data Structures and Algorithms  3 0 2 4

Unit I
Abstraction - Abstract data types; Data Representation; Elementary data types; Basic concepts of data Structures; Mathematical preliminaries - big-Oh notation; efficiency of algorithms; notion of time and space complexity; performance measures for data structures. ADT array - Computations on arrays - sorting and searching algorithms.

Unit-II
ADT Stack, Queue, list - array, linked list, cursor based implementations of linear structures. ADT Tree - tree representation, properties traversal of trees; ADT- Binary Trees – properties and algorithms .

Unit-III
ADT Priority Queue - Heaps; heap-based implementations; applications of heaps - sorting; Search Tree - Binary search tree; balanced binary search trees - AVL tree; Applications of Search Trees - TRIE; 2-3-4 tree; concept of B-Tree. ADT Dictionary - array based and tree based implementations; hashing - definition and application.

Unit-IV
Introduction to time complexity. Bio-O, worst case complexity, polynomial classifications. Satisfiability, NP Complete and NP Hard (Definitions only).

Unit-V

Text Books:

Reference Books:

20MAT508 Optimization Techniques 3 1 0 4

Unit I
Single Variable Non-Linear Unconstrained Optimization
One dimensional Optimization methods, Uni-modal function, Region elimination methods - interval halving, Fibonacci search, Golden section search, Point estimation method - successive quadratic search, Gradient based methods – Newton’s method, secant method.

Unit II
Multi Variable Non-Linear Unconstrained Optimization

Unit III
Constrained optimization

Unit IV
Integer Programming and Dynamic Programming

Unit V
Specific Search Algorithms
Hill Climbing, Simulated Annealing, Genetic Algorithms, Ant Colony Optimization.

Text Books:
20MAT509 Introduction to Data Science 3 0 2 4

Course Outcomes
CO1: Exploring and implementing exploratory data analytics
CO2: Understanding correlation and regression and visualising them using R
CO3: Understanding supervised learning through linear and logistic regressions
CO4: Understanding and implementing classifiers for unsupervised data
CO5: Exploring Massive data sets and implementing classification algorithms

Unit I
Data Collection, classification and analysis - Sampling methods, classification of data and representation of data- bar and pie charts – histogram frequency polygon – Box plot. Data Analysis Measures of Central tendency and dispersion - Mean, median, mode, absolute, quartile and standard deviations, skewness and kurtosis for both grouped and ungrouped data. Association of attributes.

Unit II
Curve fitting and interpolation - Fitting of straight lines and curves - Correlation, regression, fitting of simple linear lines, polynomials and logarithmic functions - Interpolation and extrapolation methods - Binomial expansion, Newton and Gauss methods.

Unit III
Supervised Learning (Regression/Classification): Basic methods: Distance-based methods, Nearest-Neighbors, Decision Trees, Naïve Bayes. Linear models: Linear Regression, Logistic Regression, Generalized Linear Models. Support Vector Machines,

Unit IV

Unit V

Text Books /Reference books:
20MAT510 Python Programming 3 02 4

Unit I

Unit II
Data Structures: Lists - Operations, Slicing, Methods; Tuples, Sets, Dictionaries, Sequences Comprehensions. Case Study: Nondirective Psychotherapy

Unit III

Unit IV
Data Gathering and Cleaning: Cleaning Data, Checking for Missing Values, Handling the Missing Values, Reading and Cleaning CSV Data, Merging and Integrating Data, Reading Data from the JSON Format, Reading Data from the HTML Format, and Reading Data from the XML Format.

Regular expressions: Character matching in regular expressions, Extracting data using regular expressions, Combining searching and extracting and Escape character. Case Study: Detecting the e-mail addresses in a text file.

Unit V
Popular Libraries for Data Visualization in Python: Matplotlib, Seaborn, Plotly, Geoplotlib, and Pandas. Data Visualization: Direct Plotting, Line Plot, Bar Plot, Pie Chart, Box Plot, Histogram Plot, Scatter Plot, Seaborn Plotting System, Strip Plot, Box Plot, Swarm Plot, Joint Plot, Matplotlib Plot, Line Plot Bar Chart, Histogram Plot, Scatter Plot, Stack Plot and Pie Chart.

Coding Simple GUI-Based Programs: Windows and Labels, Displaying Images, Command Buttons and Responding to Events, Viewing the Images of Playing Cards, Entry Fields for the Input and Output of Text, and Using Pop-up Dialog Boxes. Case Study: A GUI-Based ATM

Text Books:


Reference Books

1. https://www.w3schools.com/python
2. Learning Python, Mark Lutz, Orielly

SEMESTER-II

20MAT515 Statistical Inference and Design of Experiments 3 1 0 4

CO1: To understand the concept of testing of hypothesis of various parameters using single sample and apply to engineering, science and business problems.
CO2: To know to apply goodness of fit tests and nonparametric tests
CO3: To understand statistical inference for two samples and apply to engineering, science and business problems.
CO4: To develop experiments and analyse the variance to conclude on the parameters of the population involved
CO5: To construct factorial experiments and to use for various real time problems
Unit I

Unit II
Statistical Inference for Two Samples -Inference on the Difference in Means of Two Normal Distributions, Variances Known- Inference on the Difference in Means of two Normal Distributions, Variances Unknown- Paired \( t \)-Test- Inference on the Variances of Two Normal Distributions- Inference on Two Population Proportions.

Unit III
Goodness of Fit Tests and Categorical Data Analysis - Goodness of Fit Tests When all Parameters are Specified- Goodness of Fit Tests When Some Parameters are Unspecified - Tests of Independence in Contingency Tables- Tests of Independence in Contingency Tables Having Fixed Marginal Totals- The Kolmogorov–Smirnov Goodness of Fit Test for Continuous Data. Nonparametric Procedures -The Sign Test, The Wilcoxon Signed-Rank Test, Comparison to the \( t \)-Test, Equivalence Testing- The Runs Test for Randomness

Unit IV

Unit V
Design of Experiments with Several Factors-Introduction, Factorial Experiments, Two-Factor Factorial Experiments, General Factorial Experiments, \( 2^k \)Factorial Designs.

Text Books /Reference Books

20MAT516 Multivariate Statistics and Regression Analysis 31 0 4

Course Outcomes
CO1: To understand the basics of multivariate random variables and sampling distributions.
CO2: To apply multivariate techniques for classification of distributions
CO3: To understand the concept of PCA and its application in clustering analysis
CO4: To gain knowledge on simple linear regression, estimation and testing of model parameters
CO5: To gain knowledge on multiple linear and nonlinear regression and estimation of model parameters
Unit-I:
Multivariate Random variables and Distribution functions – Variance - covariance matrix – correlation - Bivariate normal distribution, Multivariate normal density and its properties - Definition of Wishart matrix and its properties, Mahalanobis Distance. Sampling distributions of $\overline{X}$ and $S$, Large sample behaviour of $\overline{X}$ and $S$.

Unit-II:
Classification for two populations, classification with two multivariate normal populations, Fisher’s discriminant functions for discriminating several population.

Unit-III:
Principal components analysis, Dimensionality reduction, Factor Analysis- factor loadings using principal component analysis, Cluster Analysis- Cluster Analysis: Hierarchical Clustering and divisive clustering methods.

Unit-IV:

Unit-V:
Multiple Linear Regression: Estimation of model parameters. Nonlinear Regression models, Examples of nonlinear regression models.

Text books/ Reference books:

20MAT517  Machine Learning  3  0  2  4

Course Objectives:
CO1: To be able to formulate machine learning problems corresponding to different applications.
CO2: To understand a range of machine learning algorithms along with their strengths and weaknesses.
CO3 : To understand the basic theory underlying machine learning.
CO4: To be able to apply machine learning algorithms to solve problems of moderate complexity.

Unit-I

Unit-II

Unit-III
Instance-Based Learning: Introduction, k-Nearest Neighbour Learning, Locally Weighted Regression, Radial Basis Functions, Case-Based Reasoning, Remarks on Lazy and Eager Learning.


Unit-IV

Unit-V
Reinforcement Learning: Introduction, the Learning Task, Q Learning, Non-Deterministic, Rewards and Actions, Temporal Difference Learning, Generalizing from Examples, Relationship to Dynamic Programming. Case Study.

Text Books:
1. Machine Learning – Tom M. Mitchell, - MGH

Reference Books:
2. Richard o. Duda, Peter E. Hart and David G. Stork, pattern classification, John Wiley & Sons

20MAT518 Big Data Analytics and Hadoop 3 0 2 4

Course Outcomes
CO1: Understanding the concepts of Big Data
CO2: Understanding the aspects of managing, cleaning and sampling of Data
CO3: Understanding Hadoop architecture and implement Map Reduce concept
CO4: To understand the aspects of Data base management and Querying system
CO5: Understanding and executing HDFS using PIG and HIVE
Unit I
Introduction to Big Data: Types of Digital Data—Characteristics of Data – Evolution of Big Data - Definition of Big Data - Challenges with Big Data - 3Vs of Big Data - Non Definitional traits of Big Data - Business Intelligence vs. Big Data - Data warehouse and Hadoop environment.

Unit II
Big Data Analytics: Classification of analytics - Data Science - Terminologies in Big Data Data science process – roles, stages in data science project – working with data from files — exploring data – managing data – cleaning and sampling for modeling and validation. working with relational databases - NoSQL: Types of Databases – Advantages – NewSQL - SQL vs. NOSQL vsNewSQL.

Unit III

Unit IV

Unit V
Hadoop Eco systems: Hive – Architecture - data type - File format – HQL – SerDe - User defined functions - Pig: Features – Anatomy - Pig on Hadoop - Pig Philosophy - Pig Latin overview - Data types - Running pig - Execution modes of Pig - HDFS commands - Relational operators - Eval Functions - Complex data type - Piggy Bank - User defined Functions - Parameter substitution - Diagnostic operator.

Text Books / Reference Books:
3. Data Science and big data analytics : Discovering, analyzing , visualizing and presenting data ,EMC Education Services,John Wiley 2015

20MAT519 Data Mining

Course Objectives:
CO1: Learn data mining basic concepts and understand association rules mining.
CO2: Capable of grouping data using clustering techniques.
CO3: Able to identify the outliers of the given dataset.
CO4: Capable of minimizing dimensionality of the data with minimum loss of information.
CO5: Able to prioritize the web links and advertisements
Unit - I
Introduction to Data Mining: Introduction, What is Data Mining, Definition, KDD, Challenges, Data Mining Tasks, Data Preprocessing, Data Cleaning, Missing data, Dimensionality Reduction, Feature Subset Selection, Discretization and Binaryzation, Data Transformation; Measures of Similarity and Dissimilarity- Basics.

Unit - II

Unit-III

Unit-IV
Outlier Detection: Outliers and Outlier Analysis -What Are Outliers?, Types of Outliers, Challenges of Outlier Detection, Outlier Detection Methods, Statistical Approaches, Parametric Methods, Nonparametric Methods, Proximity-Based Approaches, Clustering-Based Approaches, Classification-Based Approaches, Mining Contextual and Collective Outliers.

Unit-V

Text Books:

Reference Books and websites:
1. https://nptel.ac.in/courses/106/105/106105174/

20MAT520 Data Security 3 0 0 3

Unit I
Access control mechanisms in general computing systems; Authentication and authorization mechanisms- Passwords (Single vs Multifactor), Captcha, Single Sign-on- Oauth and OpenId connect, Authentication Protocols (Kerberos, X.509).
Unit II
Malwares and its protection mechanisms- Viruses, Worms, Trojans, Ransomware, Polymorphic malware, Antivirus, Firewall and Intrusion detection systems.

Unit III
Networking Basics, Web, Email, and IP Security- SSL, TLS, WEP, SET, Blockchain, PGP, IPSEC.

Unit IV

Unit V

Textbook:

References:

SEMESTER-III
20MAT606 STATISTICAL QUALITY CONTROL AND RELIABILITY 3 1 0 4
CO1 To understand the basic concepts of quality control and to construct variable and attribute control charts.
CO2 To understand and construct EWMA and CUSUM charts, analyse the process capability and Six Sigma quality metrics.
CO3 To gain knowledge about acceptance sampling methods and their properties
CO4 To gain knowledge about reliability and properties
CO5 To study reliability distributions and analyse reliability of systems and maintenance

Unit- I
Basic concept of quality control, process control and product control, Statistical process control, theory of control charts, Shewhart control charts for variables- R, s charts, attribute control charts - p, np, c, u charts, modified control charts.

Unit- II
OC and ARL curves of control charts, moving average control charts, EWMA charts, CUSUM charts, – two sided and one sided procedures – V – mask technique, process capability analysis, process capability indices – C_p and C_pk, Six Sigma quality metrics

Unit- III
Acceptance sampling for attributes, single sampling, double sampling, measuring performance of the sampling plans- OC, AQL, LTPD, AOQ, ATI curves.

Unit -IV
Introduction to Reliability and its needs; Different Approaches to Reliability Analysis, Application Areas, State Variable, Time to Failure, Failure Rate Function, Mean Time to Failure, Relationship between the Functions F(t), f(t), R(t), and z(t) , Bath tub curve, Mean time to failure, Residual time

Unit - V
Parametric families of some common life distributions –Exponential, Weibull and Gamma and its characterization-Reliability estimation of parameters in these models. Fault Tree Analysis, Reliability Block Diagrams, Systems of Independent Components -System Reliability, Nonrepairable Systems, Quantitative Fault Tree Analysis, Reliability of Maintained Systems -Types of Maintenance, Downtime and Downtime Distributions, System Availability Assessment

Text Books/References:

**20MAT607**  **Introduction to Deep Learning**  **3 0 2 4**

Course outcomes

CO1: Understand the basics concepts of artificial neural networks.
CO2: Gain knowledge about activation functions and understand the multi layer neural network
CO3: Know the importance of regularization, bagging and ensemble methods
CO4: To gain knowledge convolution neural network and case studies
CO5: To gain knowledge about recurrent neural networks, adversarial neural networks, Spectral CNN and deep reinforcement learning

Unit-I
Biological neuron, idea of computational units, McCulloch – pitts unit and thresholding logic, linear perceptron, perceptron learning algorithm, convergence theorem for Perceptron learning algorithm, logistic regression, gradient descent.

Unit-II
Feed forward neural network, activation functions, non-linear activation functions. multi-layer neural network.

Unit-III
Practical aspects of deep Learning: training, testing, regularization –dataset augmentation, Noise robustness, multitask learning, bagging and other ensemble methods, dropout- generalization.

Unit-IV

Unit-V

TextBooks/ Reference Books
Taguchi loss functions – mean square error loss function, average loss function, higher the better and lower the better loss functions – two-way analysis of variance with interactions – factorial experiments with two and three-level factors – orthogonal array experiments with two and three-level factors – methods of interpretation of experimental results - parameter and tolerance design experiments – signal-to-noise ratios – inner and outer array experiments.

**Text/Reference Books**

1. Taguchi Techniques for Quality Engineering

**20MAT652 Special Distribution Functions**


**Text/Reference books**


**20MAT653 Pattern Recognition**


Text Reference Book:


20MAT654 Stochastic Process 3 0 0 3


Text Books:


Reference Books:


20MAT655 Queuing Theory 3 0 0 3

Queuing Models: Basic characteristics of a Queueing Model – Role of Poisson and Exponential distributions, Stochastic Processes, Markov chains, Poisson Processes, Poisson Queuing Models with single server: Descriptions of the model, Assumptions, Probability distributions for number of Units (steady state), waiting time distribution, simple numerical problems on (M/M/1): (/FIFO) and (M/M/1): (N/FIFO) Models.
Poisson Queuing Models with multiple server: Descriptions of the model, Assumptions, Probability distributions for number of Units (steady state), waiting time distribution, simple numerical problems on (M/M/C): (/FIFO), (M/M/C): (N/FIFO) and (M/M/C): (C/FIFO) Models, M/M/G Models.

Text Books

Reference Books

20MAT656 Market Analytics

Business Analytics Basics: Definition of analytics, Evolution of analytics, Need of Analytics, Business analytics vs business analysis, Business intelligence vs Data Science, Data Analyst Vs Business Analyst, Business Analytics at the Strategic Level, Functional Level, Analytical Level, Data Warehouse Level. Market Segmentation Variables, Market Segmentation Types, Marketing Data Landscape, Analyzing the trend of data in Marketing—case studies.

Time series as a discrete parameter stochastic process, Auto- covariance, Auto-correlation functions and their properties. Exploratory time series analysis, Test for trend and seasonality, Exponential and moving average smoothing, forecasting based on smoothing.


Text / References Books:


20MAT657 Survival Analysis

Survival Analysis: Functions of survival times, survival distributions and their applications Censoring Schemes: Type I, Type II and progressive or random censoring with biological
examples. Estimation of mean survival time and variance of the estimator for Type I and Type II censored data with numerical examples.


Competing Risk Theory: Indices for measurement of probability of death under competing risks and their inter-relations. Estimation of probabilities of death using maximum likelihood principle and modified minimum Chi-square methods.

References
2. Collet, D. *Statistical analysis of life time data*, 1984

20MAT658 Sampling Techniques 3003

Preliminary concepts – schedules and questionnaires, pilot survey, non-sampling errors, use of random numbers. Simple random sampling with and without replacements, random number generation– estimates of population mean and population proportion and their standard errors, Probability proportional to size sampling, estimates of these standard errors. Stratified random sampling – estimates of sample statistic and estimates of their standard errors. Allocation of sample size in stratified random sampling. Linear and circular systematic sampling. Cluster sampling: Two stage sampling (equal first stage units). Ideas of ratio and regression estimators – only estimates of sample mean.

References

20MAT659 Demography and Actuarial Statistics 3003

Demographic data – Sources, Coverage and Content errors in demographic data. Measures of fertility period and cohort measures. Use of birth order Statistics and child - Woman ratio. Brass

Life table: Basic definitions, probabilities, construction of life tables, life expectancy. Life annuities: calculating annuity premium, interest and survivorship discount function, guaranteed payments, deferred annuities.

Life insurance: Introduction, calculation of life insurance premiums, types of life insurance, combined benefits, insurances viewed as annuities. Insurance and annuity reserves: General pattern reserves, recursion, detailed analysis of an insurance.

Contingent Functions: Contingent probabilities, assurances. Decrement tables. Pension funds: Capital sums on retirement and death, widow’s pensions, benefits dependent on marriage.

Text Books:


20MAT660 Official Statistics 3 0 0 3


Text Books:

2. Principles and Accommodation of National Population Census, UNEDCO

20MAT661 Healthcare Analytics 3 0 0 3

**Text / References books :**


**20MAT662 Computational Biology 3 0 0 3**


**References/ Textbooks**


**20MAT663 Computer aided drug designing 3 0 0 3**

References/ Textbooks


20MAT664 Reinforcement Learning 3 0 0 3


Text/ References Book:
An Introduction to social network data analytics: research Issues, statistical properties of social networks, random walks in social networks and their applications: survey, applications, community discovery in social networks, node classification in social networks, evolution in social networks - survey, survey of models and algorithms for social influence analysis, survey of algorithms and systems for expert location in social networks, survey of link prediction in social networks, data mining in social media, text mining in social networks

**Text and Reference**


---

**20MAT666 Mining of Massive Datasets**


**Text / References Book**


System models : physical models, architecture models, operating system support. Distributed file systems – introduction- time and global states – synchronization of physical clocks – coordination and agreements: Mutual exclusion, election, consensus.

Text Books


References