SEMESTER – I

MBD 411 STATISTICAL METHODS

Unit 1
Data Collection & Visualization: Concepts of measurement, scales of measurement, design of data collection formats with illustration, data quality and issues with data collection systems with examples from business, cleaning and treatment of missing data, principles of data visualization, and different methods of presenting data in business analytics.

Unit 2
Basic Statistics: Frequency table, histogram, measures of location, measures of spread, skewness, kurtosis, percentiles, box plot, correlation and simple linear regression, partial correlation, probability distribution as a statistics model, fitting probability distributions, empirical distributions, checking goodness of fit through plots and tests.

Unit 3
Contingency Tables: Two way contingency tables, measures of association, testing for dependence.

SUGGESTED BOOKS:
MBD 412 PROBABILITY DISTRIBUTIONS

*Prerequisite: UG Level Calculus*

**Unit 1**
Discrete Distributions: Binomial, Poisson, multinomial, hypergeometric, negative binomial, uniform. The (a,b,0) class of distributions. Moments, quantiles, cdf, survival function and other properties.

**Unit 2**
Continuous Distributions: Uniform, Normal, Exponential, gamma, Weibull, Pareto, lognormal, Laplace, Cauchy, Logistic distributions; properties and applications. Functions of random variables and their distributions using Jacobian of transformation and other tools.

**Unit 3**
Bivariate normal and bivariate exponential distributions.

**Unit 4**

**Unit 5**

**MAIN REFERENCE:**

**ADDITIONAL REFERENCES:**
MBD 413 LINEAR ALGEBRA & MATRIX THEORY

Unit 1
Fields, System of Linear Equations

Unit 2
Matrices, Elementary Row Operations, Row Reduced Matrices, Invertible Matrices, Vector spaces, Subspaces, Linear Combinations, Linear span, Linear dependence and Linear independence of vectors, Basis and Dimension, Ordered Basis, Finite dimensional vector spaces, Sum and Direct sum of subspaces,

Unit 3
Linear transformations and their representation as matrices, Kernel and Image of a linear transformation, Rank and Nullity Theorem, Change of Basis,

Unit 4
Eigen values and Eigen vectors of a linear transformation (matrices), Characteristic polynomial and minimal polynomial, Diagonalization of linear operators, invariant subspaces, Jordon form

Unit 5

SUGGESTED BOOKS:
2. Linear Algebra, Friedberg, Pearson Education.
5. Finite Dimensional Vector Spaces, P.R. Halmos, Springer-Verlag, New York.
MBD414 COMPUTING FOR DATA SCIENCES

Unit 1
Computer Package: Usage of R & Python with illustration.

Unit 2
Concepts of Computation: Algorithms, Convergence, Complexity with illustrations, some sorting & searching algorithms, some numerical methods e.g. Newton Raphson, Steepest ascent.

Unit 3
Computing Methodologies: Monte-Carlo simulations of random numbers and various statistical methods, memory handling strategies for big data.

Computer Package (20 hrs. – Theory 8 hrs. + Lab 12 hrs.)
L0 -- Installation of RStudio and understanding the basic framework
L1 – Basic computational structures – Iterations and Recursions
L2 -- Sequences and Arrays in R – Search and Sort Algorithms
L2 -- Vectors and Matrices in R – Solving systems of linear equations
L3 -- Functions in R – Plotting (2D, Contour, 3D), Differentiation, Root finding
L4 – Linear Models in R – Gradient descent, Linear regression
L5 – Eigenvalue/vector computation and SVD in R (advanced)
L6 – Handling sparse matrices in R – Basic operations on sparse matrices
L7 – Probability Distributions and Random Sampling in R
L8 – Monte-Carlo Simulation in R – Implementation of case studies

Concept of Computations (10 hrs. – Theory 6 hrs. + Lab 4 hrs.)
L1 -- Algorithms – Search and Sort, Divide and Conquer, Greedy Algorithms – motivating example from set cover for large data sets.
L2 – Computational Complexity – Growth of functions, Order notation
L3 – Computational Complexity -- Convergence, Error Estimation
L6 – Sparse Matrix – Store, Search and Basic operations
L7 – Binary Trees and Graphs as Computational Models

Numerical Methods (10 hrs. – Theory 4 hrs. + Lab 6 hrs.)
L2 -- Solving system of linear equations – Gauss-Jordan (concept of pivoting)
L3 -- Solving non-linear equations – Newton-Raphson, Steepest Descent
L4 – Optimizing cost functions – Gradient descent, Least square regression
L5 -- Iterative methods in Linear Algebra – Power iteration, Eigenvalues, SVD

Computing Methodologies (10 hrs. – Theory 4 hrs. + Lab 6 hrs.)
L7 – Random sampling from Probability Distributions, Simulated Annealing
L8 – Monte-Carlo Simulation – Case studies

Memory Handling (10 hrs. – Theory 2 hrs. + Lab 8 hrs.)
L6 – Sparse Matrix – Store, Search and Basic operations
L9 – Pruning and Sampling algorithms, Streaming data, External sorting
SUGGESTED BOOKS:
1. Software for Data Analysis – Programming with R: John M. Chambers, Springer
2. Elementary Numerical Analysis – An Algorithmic Approach: Samuel Conte and Carl de Boor (McGraw-Hill Education)
MBD 415 DATABASE MANAGEMENT

Unit 1
Basic Concepts: Need, purpose and goal of DBMS, Three tier architecture, ER Diagram, data models- Relational, Network, Hierarchical. Database Design: Conceptual data base design, concept of physical and logical databases, data abstraction and data independence, data aggregation

Unit 2
Relational data base: Relations, Relational Algebra, Theory of Normalization, Functional Dependency, Primitive and Composite data types,

Unit 3
Application Development using SQL: DDL and DML, Host Language interface, embedded SQL programming, Stored procedures and triggers and views, Constraints assertions. NoSQL Databases

Unit 4
Internal of RDBMS: Physical data organization in sequential, indexed random and hashed files. Inverted and multilist structures, B trees, B+ trees, Query Optimization, Join Algorithm, Statistics and Cost Base optimization. Parallel and distributed data base.

Unit 5
Transaction Processing, concurrency control, and recovery management. Transaction model properties and state serializability. Lock base protocols, two phase locking.

SUGGESTED BOOKS:
3. Hansen and Hansen : DBM and Design, PHI
MBD 416 PROFESSIONAL COMMUNICATION

Unit 1

Unit 2
Oral Communication: Listening Skill – Active listening, Barriers to active listening, Speaking Skill- Stress patterns in English, Questioning skills, Barriers in Speaking, Reading Skill- Skimming, Scanning, Intensive reading, linking devices in a text, Different versions of a story/ incident

Unit 3
Written communication: Writing process, paragraph organization, writing styles, Types of Writing - Technical vs. creative; Types of technical writing, Scientific Writing: Writing a Scientific Report

Unit 4

Unit 5
Presentation Skills: (i) How to make power point presentation (ii) Body language during presentation (iii) Resume writing: Cover letter, career objective, Resume writing (tailor made). Interview Skills: Stress Management, Answering skills.

SUGGESTED BOOKS:
1. Advanced English Usage: Quirk & Greenbaum; Pearson Education.
5. Communication of Business; Taylor, Shirley; Pearson Publications.
12. Wren & Martin: High School English Grammar and Composition
14. Martin Hewings: Advanced Grammar in Use
15. Betty Schrampfer: Understanding and Using English Grammar
MBD417  PYTHON AND JAVA
Prerequisite: Programming Concepts

Unit 1
Introduction to Python: The basic elements of python, Branching Programs, Control Structures, Strings and Input, Iteration, Functions, Scoping and Abstraction, Functions and scoping, Specifications, Recursion, Global variables, Modules, Files, System Functions and Parameters,

Unit 2

Unit 3
Regular Expressions – REs and Python, Plotting using PyLab

Unit 4
Introduction to Java: Overview and characteristics of Java, Java program compilation and execution process, Organization of JVM, JVM as interpreter and emulator, sandbox model. Data Types, primitive variables, arrays, operators, Control statements, standard input-output and main method. Object Oriented concepts: Concept of encapsulation and abstraction, Designing Classes, objects, instance variables and methods, Class modifiers, Inheritance, Interfaces, Abstract classes, Polymorphism: overloading and overriding, Composition.

Unit 5
Constructors: use of this and super, Java Stack and Heap, Garbage collection. Static methods and variables, Wrapper classes, Autoboxing, Standard classes: Math and String

Unit 6
Exception Handling & applications: exception types, nested try-catch, throw, throws and finally statements. Multithread Programming: thread creation, synchronization and priorities. Input-output and file operations: Java.io, Object serialization and deserialization. Java Collections API: Arraylist, Set, list, Map, Hashtable, Comparator and comparable Database connectivity, Java Packages, creating a jar file

SUGGESTED BOOKS:
SEMESTER- II

MBD 421 FOUNDATIONS OF DATA SCIENCE

Unit 1
High Dimensional Space: Properties, Law of large number, Sphere and cube in high dimension, Generation points on the surface of sphere, Gaussians in high dimension, Random projection, Applications.

Unit 2
Random Graphs: Large graphs, G(n,p) model, Giant Component, Connectivity, Cycles, Non-Uniform models, Applications.

Unit 3
Singular Value Decomposition (SVD): Best rank k approximation, Power method for computing the SVD, PCA.

Unit 4

Unit 5
Algorithm for Massive Data Problems, Frequency moments of data streams, Matrix algorithms using sampling.

Unit 6

SUGGESTED BOOKS:
1. Foundation of Data Science: Avrim Blum, John Hopcroft, and Ravindran Kannan
MBD 422 ADVANCE STATISTICAL METHODS

Unit 1
Estimation: Unbiasedness, Consistency, UMVUE, Maximum likelihood estimates.

Unit 2
Test of Hypotheses: Two types of errors, test statistic, parametric tests for equality of means & variances.

Unit 3
Gauss Markov Model, least square estimators, Analysis of variance.

Unit 4
Regression: Multiple linear regression, forward, backward & stepwise regression, Logistic regression

SUGGESTED BOOKS:
2. Introduction to Linear Regression Analysis : Douglas C. Montgomery
MBD 423 MACHINE LEARNING

Unit 1
Basics: Introduction to Machine Learning, Different Forms of Learning.

Unit 2
Regression Analysis: Linear Regression, Ridge Regression, Lasso, Bayesian Regression, Regression with Basis Functions. **Classification Methods**: Instance-Based Classification, Linear Discriminant Analysis, Logistic Regression, Large Margin Classification, Kernel Methods, Support Vector Machines, Multi-class Classification, Classification and Regression Trees.

Unit 3

Unit 4

Unit 5
Ensemble Methods: Boosting - Adaboost, Gradient Boosting, Bagging - Simple Methods, Random Forest.

Unit 6

Unit 7
Dimensionality Reduction: Principal Component Analysis, Independent Component Analysis, Multidimensional Scaling, and Manifold Learning.

**SUGGESTED BOOKS:**

**Recommended Textbooks:**

**Additional Textbooks:**
1. Pattern Classification. R.O. Duda, P.E. Hart and D.G. Stork.
2. Data Mining: Tools and Techniques. Jiawei Han and Michelline Kamber.
MBD 424 VALUE THINKING

Unit 1
This course involves watching few movies (list provided below) such as Twelve Angry Men, Roshman, and reading few books (list provided below) that deals mostly with argumentative logic, evidence, drawing inference from evidences. After watching each movie and reading each book, there will be general discussion amongst the students. Each student will prepare a term paper. Evaluation will be on the basis of this term paper and participation in group discussion.

Movies:
1. Twelve Angry Men
2. Roshoman by Kurosawa
3. Trial of Nuremberg
4. Mahabharata by Peter Brook

Books:
1. The Hound of the Baskervilles by Arthur Conan Doyle
2. Five Little Pigs by Agatha Christie
3. The Purloined Letter by Edger Allan Poe
4. The Case of the Substitute Face
MBD 425 COMBINATORIAL OPTIMIZATION

Unit 1
Linear Optimization Problem: Linear Programming, Simplex method, Revised Simplex method, Duality, Dual Simplex, Interior Point Method, Transportation problem, Assignment Problem

Unit 2
Non-linear Optimization Problem: General Non-Linear Unconstrained Optimization, convex function, concave function, local & global optimum, quadratic programming

Unit 3
Discreet Optimization Problem: Local search, Greedy Algorithm, Dynamic Programming, Branch & Bound Algorithm, Network Flow Problem: Shortest Path Problem, Knapsack problem, Max-Flow and Min-cut problem

Unit 4
Combinatorial Optimization Problems in Computer vision, social networks, cyber physical systems, Big Data analytics

SUGGESTED BOOKS:
MBD 426 INTRODUCTION TO ECONOMETRICS & FINANCE

Unit 1
Basics of Finance: Time value of money, concept of present and future value analysis, stock and bond valuations, risk and return, Systematic and unsystematic risk, Diversification, cost of capital, capital structure, Dividend Discount Model, Portfolio Theory, Efficient Market Hypothesis (EMH), Capital asset pricing model (CAPM), Market Volatility, Options.

Unit 2
Introduction to Econometrics (using finance concepts): Assumptions of Classical Linear Regression Model, Ordinary Least Squares approaches, Autocorrelation, Heteroscedasticity, Multicollinearity, Dummy Variable approaches, Distributed lag models

Unit 3
A brief Introduction to Time Series and Panel data models, Components of time series, Stationary and non-stationary time series, ARMA and ARIMA models, Static panel data models: fixed effects and random effects.

SUGGESTED BOOKS:
SEMESTER – III

MBD 511 MODELLING IN OPERATIONS MANAGEMENT

Unit 1
The course involves training the students to design, manage and improve a firm’s systems and processes. The objective is to develop skills to combines data, technology, and mathematical models to help managers make better decisions, identify new opportunities, and become more competitive. The students will undertake a rigorous set of case studies that equip them with the mindset necessary to be able to classify various operations management problems, identify the nature of the information needed to be able to address the problem, translate these problems into the appropriate statistical and/or mathematical framework and interpret the results of the models in a verbal manner.

Suggested Case studies:

- Venture analytics
- Banking analytics
- Marketing analytics
- Healthcare analytics
- Retail analytics
- Supply chain analytics
MBD 512 ENABLING TECHNOLOGIES FOR DATA SCIENCE

Unit 1
Big Data and Hadoop: Hadoop architecture, Hadoop Versioning and configuration, Single node & Multi-node Hadoop, Hadoop commands, Models in Hadoop, Hadoop daemon, Task instance, illustrations.

Unit 2
Map-Reduce: Framework, Developing Map-Reduce program, Life cycle method, Serialization, Running Map Reduce in local and pseudo-distributed mode, illustrations.

Unit 3
HIVE: Installation, data types and commands, illustration.

Unit 4
SQOOP: Installation, importing data, Exporting data, Running, illustrations

Unit 5
PIG: Installation, Schema, Commands, illustrations.

SUGGESTED BOOKS:
MBD 513 DATA MINING

Unit 1
Data Mining: Introduction, Techniques, Issues and challenges, applications, Data preprocessing, Knowledge representation

Unit 2
Association Rule Mining: Introduction, Methods to discover association rules, Association rules with item constraints

Unit 3
Decision Trees: Introduction, Tree construction principle, Decision tree construction algorithm, pruning techniques, Integration of pruning and construction; Cluster analysis: Introduction, clustering paradigms, Similarity and distance, Density, Characteristics of clustering algorithms, Center based clustering techniques, Hierarchical clustering, Density based clustering, Other clustering techniques, Scalable clustering algorithms, Cluster evaluation

Unit 4
Rough set theory, use of rough set theory for classification & feature selection.

Unit 5
ROC Curves: Introduction, ROC Space, Curves, Efficient generation of Curves, Area under ROC Curve, Averaging ROC curves, Applications

Unit 6

SUGGESTED BOOKS:
MBD 514 TIME SERIES & FORECASTING

Unit 1

Unit 2
Time Series Models: White noise Process, Random walk, MA, AR, ARMA and ARIMA models, Box- Jenkins’s Methodology fitting of AR(1), AR(2), MA(1), MA(2) and ARIMA(1,1) process. Unit root hypothesis, Co-integration, Dicky Fuller test unit root test, augmented Dickey – Fuller test.

Unit 3
Non-linear time series models, ARCH and GARCH Process, order identification, estimation and diagnostic tests and forecasting. Study of ARCH (1) properties. GARCH (Conception only) process for modelling volatility.

Unit 4

SUGGESTED BOOKS:
Main References
5. Introduction to Statistical Time Series : W.A. Fuller

Main Reading for Time Series and forecasting
MBD 515 MULTIVARIATE STATISTICS

Unit 1
Review of Multivariate Normal Distribution (MVND) and related distributional results. Random sampling from MVND, Unbiased and maximum likelihood estimators of parameters of MVND, their sampling distributions, independence. Correlation matrix and its MLE. Partial and multiple correlation coefficients, their maximum likelihood estimators (MLE), Wishart distribution and its properties (only statement).

Unit 2
Hotelling's $T^2$ and its applications. Hotelling's $T^2$ statistic as a generalization of square of Student's statistic. Distance between two populations, Mahalanobis $D^2$ statistic and its relation with Hotelling's $T^2$ statistic.

Unit 3
Classification problem – two populations, two multivariate normal populations, several populations; Discriminant analysis - Fischer's method, Logistic Regression Principle component analysis – Introduction, population principal components, summarizing sample variation by principal components, graphing principal components

Unit 4
Canonical correlation – Introduction, canonical variates & correlations, interpreting canonical variables,

Unit 5

Unit 6
Cluster Analysis – Introduction, similarity measures, hierarchical & non-hierarchical clustering methods, multidimensional scaling, correspondence analysis

SUGGESTED BOOKS:
MBD 516 CLOUD COMPUTING

Unit 1

Unit 2

Unit 3
Cloud Computing Architecture, Comparison with traditional computing architecture (client/server), Services provided at various levels, How Cloud Computing Works, Role of Networks in Cloud computing, protocols used, Role of Web services

Unit 4
Service Models (XaaS), Infrastructure as a Service (IaaS), Platform as a Service(PaaS), Software as a Service(SaaS),Deployment Models ,Public cloud, Private cloud, Hybrid cloud Community cloud, Infrastructure as a Service(IaaS). Introduction to virtualization, Different approaches to virtualization, Hypervisors, Machine Image, Virtual Machine (VM).

Unit 5
Examples, Amazon EC2, Renting, EC2 Compute Unit, Platform and Storage, pricing, customers Eucalyptus, Platform as a Service (PaaS),Introduction to PaaS, What is PaaS, Service Oriented Architecture (SOA).

Examples: Google App Engine, Microsoft Azure, SalesForce.com’s Force.com platform

Software as a Service (PaaS): Introduction to SaaS, Web services, Web 2.0, Web OS, Case Study on SaaS

Cloud Security. Case Study on Open Source & Commercial Clouds

SUGGESTED BOOKS:
MBD 517 BIOINFORMATICS

Unit 1
The objective is to train the students to learn to identify, compile, analyze, and communicate complex biological and genetic data for initiatives in areas such as human genome analysis, disease research, and drug discovery and development. The course involves study of existing algorithms and application of data analytics to biological data.

- Sequence Alignment problem & Algorithm. Pairwise & Multiple sequence Alignment.
- Advance Alignment Method.
- Gibbs Sampling
- Population Genomics.
- Genetic Mapping.
- Disease Mapping
- Gene Recognition
- Transcriptome & Evolution.
- Protein Structure.
- Protein Motifs.
- Hidden Markov Models
- Lattice Model.
- Algorithms.

SUGGESTED BOOKS:
1. Introduction Computational Molecular Biology : C Setubal & J Meidanis, PWS Publishing Boston, 1997
MBD 518 SOFTWARE ENGINEERING

Unit 1

Unit 2

Unit 3

Unit 4
UML 2.0 diagrams: Structure diagrams, Behavior diagrams

Unit 5

Unit 6

SUGGESTED TEXTBOOKS:
ADDITIONAL TEXTBOOKS:
SEMESTER – IV

MBD 521 Internship based project

A real-life project must be undertaken at an industry for 20 weeks. Each student will have two supervisors: Once from academic institution and one from the industry. The project shall involve handling data extensively and use of methodologies learnt during the course work to derive meaningful inferences. A final project report has to be submitted and an “open” presentation has to be made.

Project evaluation may be as follows.

Project evaluation of total 200 Marks
Internal Evaluation (by 3-member including internal supervisor) – 100 Marks
External Evaluation – 100 Marks