ST. XAVIER'S COLLEGE, MUMBAI

(An autonomous college affiliated to the Mumbai University)

Syllabus for the M.Sc in Big Data Analytics

(Credit Based Semester and Grading system effect from the academic year 2019-20)

Pre-course Work

1. Microsoft Excel for Data Analysis
    a. Excel Tables, Filters, Sorting
    b. Pivot Tables and Charts
    c. Formats, Formulas, Dates
    d. Functions – Mathematical, Statistical, Text, Date

    Reference:

    On-line courses/Tutorials:

    i. Microsoft Virtual Academy:
        a. Analyzing and Visualizing Data with Excel https://mva.microsoft.com/en-US/training-courses/analyzing-and-visualizing-data-with-excel-11157
        b. Data Analysis with Excel https://mva.microsoft.com/en-US/training-courses/data-analysis-with-excel-16654

    ii. Edx.Org:

    e. Introduction to Data Analysis using Excel https://www.edx.org/course/introduction-to-data-analysis-using-excel-0

    iii. Coursera.org:

    f. Introduction to Data Analysis Using Excel https://www.coursera.org/learn/excel-data-analysis
2. Basic Unix Programming
    a. Basic Unix Commands
    b. Handling files and folders
    c. Concatenation, find and replace, modify file & texts
    d. Basic summary commands

    Reference:

    On-line courses/Tutorials:

    Data Camp:

    e. Introduction to Shell for Data Science https://www.datacamp.com/courses/introduction-to-shell-for-data-science

    Linux.Org:

    f. Linux Beginner Tutorials https://www.linux.org/forums/linux-beginner-tutorials.123/
    g. Github - Organizing with Unix: https://rafaLabgithub.io/dsbook/organizing-with-unix.html

    Book:

    h. Data Science at the Command Line, JeroenJanssens, https://www.datascienceatthecommandline.com/

SEMESTER – I

All courses compulsory

1. Statistical Methods
2. Probability & Stochastic Process
3. Linear Algebra & Linear Programming
4. Computing for Data Sciences using R, Python and Java
5. Database Management – Relational and Non-Relational
6. Python Programming (Lab work is Must) - Introduction to Python interpreter, Control statements, functions, I/O, File handling, Packages/Libraries, Exception Handling, OO Programming.

SEMESTER – II

Compulsory courses:

1. Foundations of Data Science
2. Advanced Statistical Methods
3 Machine Learning I
4. Enabling Technologies for Data Science I (Lab work is must) : NoSQL database, Oozie
5. Value Thinking

Elective courses: (choose 1)

1.Operations Research

2. Cloud Computing: Introduction to Cloud computing, Cloud service methods, IaaS, PaaS, SaaS, fundamentals of cloud Architecture (load distribution, resource pooling, scalability,

3. load balancing, redundancy, etc), Introduce DevOps, CICD. Hands-on practice either on AWS/Azure/Google Cloud Platform.

4. Natural Language Processing
5. Unix Programming
6. Operating Systems
7 Multivariate Statistics

SEMESTER – III

Compulsory courses:

1. Enabling Technologies for Data Science 2 (Lab is Must):Spark, Scala, Mahout

2. Machine Learning 2 including Deep Learning

3A. Data Visualization with Tableau: Learn about design principles, human perception and effective story telling with data, dashboards, modern visualization tools and techniques (cover Tableau). Hands-on practice on Tableau is must.

3B. (30hrs.). Modelling in Operations Management – Banking Analytics, Healthcare analytics, Retail analytics

Elective courses: (choose 2)

1. Introduction to Econometrics & Finance
2. Time series Analysis & Forecasting
3. Bio informatics
4. Big Data Technologies and Architecture
5. IPR
6. Cyber Security
7. Text mining
8. Advanced Analytics

SEMESTER – IV

1. Internship based project.

Semester I

Course code	Course name	credits	Lectures
ITS0701	Statistical Method	3	45
ITS0702	Probability & Stochastic Process	3	45
ITS0703	Linear Algebra & Linear Programming	3	45
ITS0704	Computing for Data Sciences using R, Python and Java	3	45
ITS0705	Database Management – Relational and Non-Relational	3	45
ITS07PR1	Linear Algebra & Linear Programming & Statistics	3	60
ITS07PR2	Data base Management Systems and computing for data science	3	60
ITS07PR3	Python Programming	3	60
	Total	24	405

Total 24 credits

Semester II

Course code	Course name	credits	Lectures
ITS0801	Foundations of Data Science	3	45
ITS0802	Advanced Statistical Methods	3	45
ITS0803	Machine Learning-I	3	45
ITS0804	Value Thinking	2	30
ITS0805	Enabling Technologies for Data Science	3	45
ITS0806	Cloud computing /Operation Research	4	60
ITS08PR1	Foundation of Data Science & Advanced Statistical Methods	3	60
ITS08PR2	Machine Learning I &Enabling Technologies for Data Science	3	60
Total		24	390

Semester III

Course code	Course name	credits	Lectures
ITS0901	Enabling Technologies for Data Science 2	4	60
ITS0902	Machine Learning 2 including Deep Learning	4	60
ITS0903	Data Visualization with Tableau & Modelling in Operations Management	3	45
ITS0904	IPR / Cyber Security. Text mining Advanced Analytics	4	60
ITS0905	Time series Analysis & Forecasting /Bio informatics / Big Data Technologies and Architecture	4	60
ITS09PR1	Machine Learning 2 including Deep Learning&elective 1	3	60
ITS09PR2	Enabling Technologies for Data Science 2&elective 2	3	60
ITS09PR3	Data Visualization with Tableau & Modelling in Operations Management	3	60
Total		28	465

Semester IV

ITS10PJ **Internship based Project** **20 credits**

1.BASIC STATISTICAL METHODS: ITS0701 45 lect

a) **Data Collection & Visualization** : (17)
Concepts of measurement, scales of measurement, design of data collection formats with illustration, data quality and issues with date 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.

b) **Basic Statistics:** (17)
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.

c) **Contingency Tables:** (11)
Two way contingency tables, measures of association, testing for dependence.

REFERENCE BOOKS:

1. Statistics: David Freedman, Pobert Pisani & Roger Purves, WW.Norten& Co. 4th Edition 2007.

2. The visual display of Quantitative Information: Edward Tufte, Graphics Press, 2001.

3. Best Practices in Data Cleaning: Jason W. Osborne, Sage Publications 2012.

Evaluation : Theory : CIA (40%) and End semester (60%)

2. PROBABILITY & STOCHASTIC PROCESS: ITS0702 [45 Lect]

a). Basic Probability : (18)

Concepts of experiments, Outcomes, Sample space, Events, Combinatorial probability, Birthday paradox, Principle of inclusion & exclusion, Conditional probability, Independence, Bayes Theorem.

b) **Probability Distribution:** (16)

Random Variables: discrete and continuous probability models, some probability distributions: Binomial, Poisson, Geometric, Hypergeometric, Normal, exponential, Chi-square, expectation, variance and other properties of the distribution.

c) **Stochastic Process:** (5)

Markov Chains, Classification of states, Stationery distribution, limit theorems, Poisson process, illustrations and applications.

d) **Introduction to Time Series**: (6)

Components of time series, Smoothing auto correlation, stationarity, concepts of AR, MA, ARMA & ARIMA models with illustrations.

REFERENCE BOOKS:

1. A First Course in Probability: Shelden M. Ross, 2014.

2. Introduction to Stochastic Process : Paul G. Hoel, Sydney C. Port & Charles J. Stone, Waveland Press, 1987.

3. Time Series Analysis and Its Applications: Robert H. Shumway and David S. Stoffer, Springer 2010.

Evaluation:Theory : CIA (40%) and End semester (60%)

3.LINEAR ALGEBRA & LINEAR PROGRAMMING ITS0703 [45 Lect]

a) **Linear Algebra:** (30)

Linear equations and matrices, matrix operations, solving system of linear equations, Gauss-Jordan method, Concept & Computation of determinant and inverse of matrix, Eigen values and eigen vectors, Illustrations of the methods, Positive semi definite and position definite matrices, illustrations

b) **Linear Programming** (15)

Definition of the problem, convex sets, corner points, feasibility, basic feasible solutions, Simplex method

REFERENCEBOOKS:

1. Linear Algebra and Its Application: Gilbert Strang, 4th Edition, Academic Press.
Hands-On Matrix Algebra Using R (Active and Motivated Learning with Applications), Hrishikesh D Vinod, World Scientific
2. Linear Programming: G. Hadley, Addison-Wesley.

Evaluation: Theory : CIA (40%) and End semester (60%)

4.COMPUTING FOR DATA SCIENCES ITS0704 [45 Lect]

a) **Computer** Packages – R and Python**:** (10)
Usage of R and Python – data handling, data analysis, statistical modeling with illustration in python and R (10 hrs. R + 10 hrs. Python).

b) Data Structure &**Concepts of Computation** using Java (20)
Algorithms, Convergence, Complexity with illustrations, some sorting & searching algorithms, some numerical methods e.g. Newton-Raphson, Steepest ascent using Java

c) **Computing Methodologies:** (15)
Monte-Carlo simulations of random numbers and various statistical methods, memory handling strategies for big data.

REFERENCE BOOKS:

Introduction to Data Science (Data Analysis and Prediction Algorithms with R), Rafael A. Irizarry, https://rafaLabgithub.io/dsbook/

Hands-On Programming with R - Write Your Own Functions and Simulations, Grolemund Garrett, O'Reilly

Data Structures and Algorithm using Java, 6th Ed. Michael T. Goodrich and Roberto Tamassia, John Wiley & Sons, Inc

Python Data Science Handbook - Essential Tools for Working with Data, Jake VanderPlas, O'Reilly

Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, WES MCKINNEY, O'Reilly

Evaluation:Theory : CIA (40%) and End semester (60%)

5. DATABASE MANAGEMENT : ITS0705 [45Lect]

a) **Basic Concepts:** (10)
Different data models, ER and EER diagram, schema, table, Big Data Concepts and Hadoop Ecosystem

b) Relational and Non-Relational Databases**:** **(**25)
Structure, various operations, normalization, SQL, No-SQL, Graph Database, Parallel and distributed data base, Map-Reduce.
Lab using SQL/Oracle/MySql for Relational databases; Hadoop(any), MangoDB, GraphDB for Big Data

c) **Implementation:** **(**10)
ORACLE SQL/MS SQL/MySQL, Hadoop Ecosystem, Concept of database security.

SUGGESTED BOOKS

1. Database system concepts : Abraham Silberschartz, Henry F. Korth and S. Surarshan, McGraw Hill, 2011.
2. Hadoop 2 Quick-Start Guide: Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem, Douglas Eadline, Addison-Wesley, Pearson Education India; First edition (1 March 2016)
3. Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, EMC Education Services, 2015

Evaluation: **Theory : CIA (40%) and End semester (60%)**

6.Linear Algebra & Linear Programming & Statistics ITS07PR1 [60 Lect]

1. Data Collection & Visualization :	(8)
2. Basic Statistics:	(8)
3.Contingency Tables:	(4)
4. Basic Probability :	(4)
5.Probability Distribution:	(4)
6.Stochastic Process:	(6)
7.Introduction to Time Series:	(6)
8.Linear Algebra	(12)
9.Linear programming	(8)

7. Data base Management Systems and computing for data science ITS07PR2

[60 Lect]

1.	Relational Databases using MS SQL	(15)
2.	Non relational databases	(5)
3.	Implementations	(10)
4.	Computer packages :R and Python	(10)
5.	Data Structure & Concepts of Computation using Java	(10)
6.	Computing Methodologies	(10)

8. Python Programming (Lab work is Must)ITS07PR3 [60 Lect]

1.Introduction to Python interpreter	(8)
2.Control statements	(8)
3.functions &I/O	(8)
4.File handling	(8)
5. Packages/Libraries	(8)
6.Exception Handling	(8)
7.OO Programming.	(8)

SEMESTER – II

Course code	Course name	credits	Lectures
ITS0801	Foundations of Data Science	3	45
ITS0802	Advanced Statistical Methods	3	45
ITS0803	Machine Learning-I	3	45
ITS0804	Value Thinking	2	30
ITS0805	Enabling Technologies for Data Science	3	45
ITS0806	Cloud computing /Operation Research	4	60
ITS08PR1	Foundation of Data Science & Advanced Statistical Methods	3	60
ITS08PR2	Machine Learning I &Enabling Technologies for Data Science	3	60
Total		24	390

1. FOUNDATIONS OF DATA SCIENCE (PROGRAMMING FOR BIG DATA) ITS0801

45lects:

a) **Graph Theory :** (7)
Basic Concepts, Algorithms for connectedness, Shortest path, Minimum Sampling Tree,

b) **High Dimensional Space:** (11)
Properties, Law of large numbers, Sphere and cube in high dimension, Generating points on the surface of a sphere, Gaussians in High dimension, Random projection, Applications.

c) **Random Graphs :** (11)
Large graphs, G(n,p) model, Giant Component, Connectivity, Cycles, Non-Uniform models, Applications.

d) **Singular Value Decomposition (SVD):** (2)
Best rank k approximation, Power method for computing the SVD, Applications.

e) **Random Walks:** **(**2)
Reflection Principle, Long leads, Changes of Sign, Illustrations.

f) **Algorithm for Massive Data Problems:** (12)
Frequency Moments of data streams, matrix algorithms.

SUGGESTED BOOK:

1. Foundations of Data Science: John Hopcroft & Ravindran Kannan.

Evaluation : Theory : CIA (40%) and End semester (60%)

2. ADVANCED STATISTICAL METHODS : ITS0802 [45 lectures]

a) **Estimation:**(13)
Unbiasedness, Consistency, UMVUE, Maximum likelihood estimates. (15 hrs.)

b) **Test of Hypotheses:**(13)
Two types of errors, test statistic, parametric tests for equality of means & variances.

c) **Linear Model:**(9)
Gauss Markov Model, least square estimators, Analysis of variance.

d) **Regression:**(10)
Multiple linear regression, forward, backward & stepwise regression, Logistic Regression.

SUGGESTED BOOKS:

1. Statistical Inference: P. J. Bickel and K. A. Docksum, 2nd Edition, Prentice Hall.

2. Introduction to Linear Regression Analysis: Douglas C. Montgomery

Evaluation: Theory : CIA (40%) and End semester (60%)

3. MACHINE LEARNING 1 ITS0803 [45 lectures]

a) **Linear Regression** (8)
Linear Regression with Multiple variables, applications.

b) **Logistic Regression:**(6)
Model, Classification, Problem of over-fitting, Applications.

c) **Neural Networks:**(7)
Representation Learning, Different Models like single and multi-layer perceptron, back propagation, Application.

d) **Machine Learning System Design:**(3)
Evaluating a learning algorithms, handling skewed data, using large data sets.

e) **Support Vector Machines:**(4)
Model, Large Margin Classification, Kernels, SVMs in practice.

f) Unsupervised Learning. **(**5)

g) Dimensionality Reduction. (8)

h) Anomaly Detection. (4)

SUGGESTED BOOKS:

1. Machine Learning: Tom Mit chell

Evaluation:Theory : CIA (40%) and End semester (60%)

4. Enabling Technologies for Data Science I ITS0803 [45 lectures]

a) **Introduction:** (5)
Knowledge discovery from databases, scalability issues.

b) Data Warehousing: (5)
General principles, modeling, design, implementation and optimization, Cloud Computing, OLAP.

c) **Data Preparation:(**5)
Pre-processing, sub-sampling, feature selection.

d) **Classification and Prediction:** (16)
Bayes learning, decision trees, CART, neural learning, support vector machines, associations, dependence analysis, rule generation.

e) **Cluster Analysis and Deviation Detection: (**8)
Partitioning algorithms, Density bases algorithm, Grid based algorithm, Graph theoretic clustering.

f) Temporal and spatial data mining. (6)

SUGGESTED BOOKS

1. Data Mining Techniques: A. K. Pujari, Sangam Books Ltd., 2001

2. Mastering Data Mining: M. Berry and G. Linoff, John Wiley & Sons., 2000
3. Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management, Olivia Parr Rud, Wiley

Evaluation: Theory : CIA (40%) and End semester (60%)

5.VALUE THINKING : ITS0804 [45 lectures]

(It does not need full class contact hours. Students have to learn themselves through Movies and Books).

This course involves watching few movies (list provided below) and reading few books (list provided below) that deals mostly with argumentative logic, evidence, drawing inference from evidences. After watching the movies and reading the books, there will be general discussion amongst the students. Couple of case studies that involve mostly logical thinking will also be presented. Each student will prepare a term paper. Evaluation will be on the basis of this term paper and participation in group discussion.

Movies:

Twelve Angry Men
Roshoman by Kurosawa
Trial of Nuremberg
Mahabharata by Peter Brook

Books:

The Hound of the Baskervilles by Arthur Conan Doyle
Five Little Pigs by Agatha Christie
The Purloined Letter by Edger Allan Poe
The Case of the Substitute Face

Evaluation: Case Studies

Electives (Choose 1)

1. OPERATIONS RESEARCH: (60 hrs.)

a) Review of Linear Programming. (5 hrs. – Theory 5 hrs.)

b) Non-Linear Programming. (10 hrs. – Theory 6 hrs. + Lab 4 hrs.)

c) Assignment Models. (5 hrs. – Theory 1 hrs. + Lab 4 hrs.)

d) Transportation Models. (15hrs. – Theory 11 hrs. + Lab 4 hrs.)

e) Queuing Models: Characteristics of Queuing Process, Poisson Process, Birth-Death Process, Single-Server Queues, Multi-ServerQueues, Queues with Truncation, Finite-Source Queues, Numerical Techniques & Simulation.

(25 hrs. – Theory 19 hrs. + Lab 6 hrs.)

SUGGESTED BOOKS:

1. Operations Research:PREM Kumar Gupta & D. S. Hira

2. Fundamentals of Queuing Theory: Donald Gross, John F. Shortle, James M. Thompson & Carl M. Harris, Fourth Edition, Wiley

Evaluation : Theory : 70% + Practical/Lab : 30%

2. Cloud Computing: Introduction to Cloud computing, Cloud service methods, IaaS, PaaS, SaaS, fundamentals of cloud Architecture (load distribution, resource pooling, scalability, load balancing, redundancy, etc), Introduce DevOps, CICD. Hands-on practice either on AWS/Azure/Google Cloud Platform (Course content to be developed)

3. Natural Language Processing (Course content to be developed)

4. Unix Programming (Course content to be developed)

5. Operating Systems (Course content to be developed)

6. MULTIVARIATE STATISTICS : (60 hrs.)

a) Representation of multivariate data, bivariate and multivariate distributions, multinomial distribution, multivariate normal distribution, sample mean & sample dispersion matrix, concepts of location & depth in multivariate data.
(20 hrs. – Theory 12 hrs. + Lab 8 hrs.)

b) Principal Component Analysis (10 hrs. – Theory 6 hrs. + Lab 4 hrs.)

c) Classification (10 hrs. – Theory 6 hrs. + Lab 4 hrs.)

d) Factor Analysis (10 hrs. – Theory 6 hrs. + Lab 4 hrs.)

e) Clustering (10 hrs. – Theory 6 hrs. + Lab 4 hrs.)

SUGGESTED BOOKS:

1. Applied Multivariate Statistical Analysis: Richard A. Johnson and Dean W. Wichern, Prentice Hall, 2002

Evaluation: Theory: 60% + Practical/Lab : 40%

Foundation of Data Science & Advanced Statistical Methods ITS08PR1

a) **Graph Theory :** (6)

Basic Concepts, Algorithms for connectedness, Shortest path, Minimum Sampling Tree, Lab: Graph Databases, Java/Python Programming

b) **High Dimensional Space:** (6 .)

Properties, Law of large numbers, Sphere and cube in high dimension, Generating points on the surface of a sphere, Gaussians in High dimension, Random projection, Applications. Lab: Graph Databases, Java/Python Programming

c) **Random Graphs :** (6 .)

Large graphs, G(n,p) model, Giant Component, Connectivity, Cycles, Non-Uniform models, Applications.
Lab: Graph Databases, Java/Python Programming

d) **Singular Value Decomposition (SVD):** (2)

Best rank k approximation, Power method for computing the SVD, Applications.
Lab: R and Python Programming (Optional: Matlab/Octave)

e) **Random Walks:(** 6)

Reflection Principle, Long leads, Changes of Sign, Illustrations.
Lab: R and Python Programming

f) **Algorithm for Massive Data Problems:** (10)

Frequency Moments of data streams, matrix algorithms.
Lab: R and Python Programming

g) **Estimation:** (4.)

Unbiasedness, Consistency, UMVUE, Maximum likelihood estimates. (15 hrs.)

h) **Test of Hypotheses:** (4)

Two types of errors, test statistic, parametric tests for equality of means & variances.

i) **Linear Model:** (6 .)

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

j) **Regression:** (6)

Multiple linear regression, forward, backward & stepwise regression, Logistic Regression.

Enabling Technologies for Data Science & Machine learning-I ITS08PR2

a) Data Warehousing: (6 .)
General principles, modeling, design, implementation and optimization, Cloud Computing, OLAP.

b) **Data Preparation: (**4)
Pre-processing, sub-sampling, feature selection.

c) **Classification and Prediction:** (8)
Bayes learning, decision trees, CART, neural learning, support vector machines, associations, dependence analysis, rule generation.

d) **Cluster Analysis and Deviation Detection: (** 8)
Partitioning algorithms, Density bases algorithm, Grid based algorithm, Graph theoretic clustering.

e) Temporal and spatial data mining. (4)

f) **Linear Regression** (4)
Linear Regression with Multiple variables, applications.

g) **Logistic Regression:**(6)
Model, Classification, Problem of over-fitting, Applications.

h) **Neural Networks:** (6)
Representation Learning, Different Models like single and multi-layer perceptron, back propagation, Application.

i) **Machine Learning System Design:** (6)
Evaluating a learning algorithms, handling skewed data, using large data sets. (5 hrs.)

j) **Support Vector Machines:** (2)
Model, Large Margin Classification, Kernels, SVMs in practice. (5 hrs.)

k) Unsupervised Learning. **(**2)

l) Dimensionality Reduction. (2)

m) Anomaly Detection. (2)

SEMESTER – III

Enabling Technologies for Data Science 2 ITS 0901 (60 Hours) :

Spark, Scala, Mahout.

2. Machine Learning 2 including Deep Learning ITS0903 (60 Hours)

Decision Tree Classification: (6 hrs. – Theory 3 hrs. + Lab 3 hrs.)
Entropy, Gini index, Algorithms, Regression Trees.

Probabilistic Classifiers: (6 hrs. – Theory 3 hrs. + Lab 3 hrs.)
Generative and Conditional classifiers.

Hyper plane classifiers: (6 hrs. – Theory 3 hrs. + Lab 3 hrs.)
Loss functions, stochastic gradient algorithms, Perceptron algorithms.

Application of to Pattern Recognition Problems. (6 hrs. – Theory 3 hrs. + Lab 3 hrs.)

Clustering: (6 hrs. – Theory 3 hrs. + Lab 3 hrs.)
Performance criteria, K-means clustering, EM algorithm

a) Collaborative filtering (6 hrs. – Theory 3 hrs. + Lab 3 hrs.)

b) Combining models (6 hrs. – Theory 3 hrs. + Lab 3 hrs.)

c) Probabilistic graphical models (6 hrs. – Theory 3 hrs. + Lab 3 hrs.)

d) Large Scale Machine Learning: 6 hrs. – Theory 3 hrs. + Lab 3 hrs.)
Gradient descent with large data sets

e) Genetic Algorithm. (6 hrs. – Theory 3 hrs. + Lab 3 hrs.)

SUGGESTED BOOKS

1. Machine Learning: Tom Mitchell

Evaluation: Theory: 50% + Practical/Lab : 50%

). Data Visualization AND MODELLING IN OPERATIONS MANAGEMENT ITS0903

3A. Data Visualization with Tableau: (30 Hours)
Learn about design principles, human perception and effective story telling with data, dashboards, modern visualization tools and techniques (cover Tableau). Hands-on practice on Tableau is must.

3B.MODELLING IN OPERATIONS MANAGEMENT: (30 hrs.)

a) Banking analytics (10 hrs.– Lab 10 hrs.)
b) Healthcare analytics (10 hrs. – Lab 10 hrs.)
c) Retail analytics (10 hrs. – Lab 10 hrs.)

SUGGESTED BOOKS: NONE

Evaluation: Practical / Lab / Report : 100%

Electives (Choose 2)

1. INTRODUCTION TO ECONOMETRICS & FINANCE (60 Hrs.):

a) Analysis of Panel Data. (19 hrs. – Theory 16 hrs. + Lab 3 hrs.)

b) Generalized Method of Moments (GMM). (18 hrs. – Theory 16 hrs. + Lab 3 hrs.)

c) Simultaneous Equations System: (7 hrs. – Theory 4 hrs. + Lab 3 hrs.)
Least Squares, Bias Problem, Estimation Method.

d) Cointegration: (8 hrs. – Theory 2 hrs. + Lab 6 hrs.)
Concept, two variable model, Engle-Granger Method, Vector autoregressions (VAR), Vector error correlation model (VECM).

e) ARCH/GARCH/SV models, some important generalizations like EGARCH & GJR models, ARCH –M models. (8 hrs. – Theory 2 hrs. + Lab 6 hrs.)

SUGGESTED BOOKS:

1. The Econometrics of Financial Markets: J. Campbell, A.Lo and C. Mackinlay

2. Econometric Analysis: William H. Greene

Evaluation : Theory : 70% + Practical/Lab : 30%

2. TIME SERIES & FORECASTING: (60 hrs.)

a) Exploratory Analysis of Time Series. (10 hrs. – Theory 4 hrs. + Lab 6 hrs.)

b) Stationary and Non-Stationary Time Series. (5 hrs. – Theory 5 hrs.)

c) AR, MA, ARMA, ARIMA models, their properties, estimation of parameters. (20 hrs. – Theory 16 hrs. + Lab 4 hrs.)

d) Tests of Non-Stationarity – Unit Root tests. (5 hrs. – Theory 3 hrs. + Lab 2 hrs.)

e) Forecasting, Smoothing, Minimum MSE Forecast, Forecast Error. (10 hrs. – Theory 8 hrs. + Lab 2 hrs.)

f) Modelling Seasonal Time Series. (5 hrs. – Theory 3 hrs. + Lab 2 hrs.)

g) Missing Data Problem in Time Series. (5 hrs. – Theory 3 hrs. + Lab 2 hrs.)

SUGGESTED BOOKS

1. Introduction to Statistical Time Series: W. A. Fuller

2. Introduction to Time Series Analysis: P. J. Brockwell and R. A. Davis

Evaluation:Theory: 70% + Practical/Lab: 30%

3. BIOINFORMATICS (60 hrs.):

a) Sequence Alignments. (4 hrs. – Theory 2 hrs. + Lab 2 hrs.)

b) Advance Alignment Methods. (4 hrs. – Theory 2 hrs. + Lab 2 hrs.)

c) Gibbs Sampling. (8 hrs. – Theory 2 hrs. + Lab 6 hrs.)

d) Population Genomics. (4 hrs. – Theory 2 hrs. + Lab 2 hrs.)

e) Genetic Mapping. (4 hrs. – Theory 2 hrs. + Lab 2 hrs.)

f) Disease Mapping. (4 hrs. – Theory 2 hrs. + Lab 2 hrs.)

g) Gene Recognition. (4 hrs. – Theory 2 hrs. + Lab 2 hrs.)

h) Transcriptome & Evolution. (4 hrs. – Theory 2 hrs. + Lab 2 hrs.)

i) Protein Structure. (4 hrs. – Theory 2 hrs. + Lab 2 hrs.)

j) Protein Motifs. (4 hrs. – Theory 2 hrs. + Lab 2 hrs.)

k) Hidden Markov Model. (4 hrs. – Theory 2 hrs. + Lab 2 hrs.)

l) Lattice Model. (4 hrs. – Theory 2 hrs. + Lab 2 hrs.)

m) Algorithms. (8 hrs. – Theory 6 hrs. + Lab 2 hrs.)

SUGGESTED BOOKS

Introduction to Computational Molecular Biology: C. Setubal & J. Meidanis, PWS Publishing, Boston, 1997

Evaluation:Theory: 50% + Practical/Lab: 50%

4. Big Data Technologies and Architecture (Course content to be developed)

5. IPR (Course content to be developed)

6. Cyber Security (Course content to be developed)

SEMESTER – IV:

Internship based project

A real life project has to be undertaken at an industry for 20 weeks. Each student will have two supervisors: one 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.

Report from two supervisors: 200 marks (100 each)
Project report: 200 marks
Presentation: 100 marks.
Total: 500 marks