MAR ATHANASIUS COLLEGE (AUTONOMOUS)
KOTHAMANGALAM, KERALA - 686666

College with Potential for Excellence
NAAC Accredited and Reaccredited 'A+ Grade' Institution

Email: mac@macollege.in
www.macollege.in

MASTER OF SCIENCE
IN
DATA ANALYTICS

PROGRAMME STRUCTURE AND SYLLABUS

(2020 Admission onwards)

(UNDER REGULATIONS OF THE POSTGRADUATE PROGRAMMES
UNDER CREDIT SEMESTER SYSTEM-MAC-PG-CSS2020)

Contents

PREFACE

Data Analytics is a multi-disciplinary field that utilizes logical techniques, procedures, calculations and frameworks to separate information and bits of knowledge from organized and unstructured data. Data Analytics is an idea to bind together Statistics, data investigation, AI and their related techniques so as to understand and analyse the actual phenomena within the data. It utilizes procedures and hypotheses drawn from numerous fields inside the setting of Mathematics, Statistics, Computer Science and Information science. Turing award winner Jim Gray envisioned Data Science as a "fourth paradigm" of science. Data Analytics, as a field of study is making a boom all over the world. It is being applied on every aspects of life. Newer domains of applications are being found out. All these points out to the importance of getting knowledge in Data Analytics concepts for a job aspirant student.

The Board of Studies in Statistics was entrusted with the job of discussing and framing the syllabus and curriculum for starting a new P.G course in Data Analytics. The Board proceeded with the task, as per the terms of reference and guidelines given by the university in line with the proposals put forward by the University Grant Commission. The Board of studies prepared a comprehensive plan of action for introducing the new PG programme with effect from the academic year 2020-21. The proposal has been made after rigorous brainstorming sessions of the Board of Studies members. The Board of Studies observed that it is high time for the introduction of the proposed course. The syllabus is carefully formulated to meet the requirements of the industry. It is envisaged that students will have maximum opportunity to pursue careers in Data Analytics domains. We acknowledge the assistance and guidance received from the management, members and special invitees in the Board of Studies in Statistics and the university and all those who have contributed in different ways in the venture. It is recommended to keep the syllabi and curriculum up to date with periodic revisions of the same. We hope this syllabi and curriculum would enrich and equip the students to meet future challenges.

Board of Studies Members in Statistics

SL NO	NAME	OFFICIAL ADDRESS
	CHAIRMAN	
1	SUDHA V.	Assistant Professor and Head Department of Statistics Mar Athanasius College, Kothamangalam
	MEMBERS	
2	Dr. NIDHI P RAMESH	Assistant Professor Department of Statistics Mar Athanasius College, Kothamangalam
	EXPERTS (2)	
3	Dr. RANI SEBASTIAN	Assistant Professor Department of Statistics St Thomas College, Thrissur
4	Dr. G. RAJESH	Associate Professor Department of Statistics Cochin University of Science & Technology Kalamassery, Kochi.
	UNIVERSITY NOMINEE	
5	Dr. JAMES KURIEN	Associate Professor and Head. Department of Statistics, Maharajas College, Ernakulam
	MEMBER FROM INDUSTRY	
6	Dr. D. DHANURAJ	Chairman, Centre for Public Policy Research, Kochi
	MERITORIOUS ALUMNUS	
7	Dr. ABDUL SATHAR E. I.	Professor, Department of Statistics, Kerala University, Kariavattom
	SPECIAL INVITEES	
8	Dr. T. M. JACOB	Associate Professor Department of Statistics Nirmala College, Muvattupuzha

9	Dr. MADHU S NAIR	Associate Professor Department of Computer Science Cochin University of Science& Technology Cochin-22
10	Mr. CENCYMON M. A.	Data Scientist, Genpact, Germany
11	Mr. MAHESH DIVAKARAN	Statistical Programmer, Genpro Research, Trivandrum

Curriculum for M.Sc. Data Analytics Programme

The Board of Studies in Statistics proceeded with the task of framing a new postgraduate course in Data Analytics in Mar Athanasius College (Autonomous) as per the terms of reference and guidelines given by the University and Kerala State Higher Education Council.

The Board of Studies resolved to formulate the curriculum and syllabi of M.Sc. course in Data Analytics, under the newly proposed credit and semester system. Programme models proposed by the M.G. University and the Kerala State Higher Education Council are selected as the base for the task. The formulation is attempted in such a way as to lay emphasis on student choice, industry requirements and self-learning.

Since all the programmes within the same stream should have the same number of credits, we have chosen 80 credits as instructed. Total number of courses in M.Sc. Data Analytics programme is stipulated as 20 which is spread over four semesters.

Introduction

Data Analytics is a multi-disciplinary field that utilizes logical techniques, procedures, calculations and frameworks to separate information and bits of knowledge from organized and unstructured data. Data Analytics is an idea to bind together Statistics, data investigation, Artificial Intelligence and their related techniques so as to understand and analyse the actual phenomena within the data. It utilizes procedures and hypotheses drawn from numerous fields inside the setting of Mathematics, Statistics, Computer Science and Information science.

M.Sc. Data Analytics is a Postgraduate degree course offering a high-level training in Data Science concepts like Machine Learning, Artificial Intelligence, etc. The program covers industry-relevant and applied topics from Mathematics, Statistics, Information Science and Computer Science. Training is delivered through application focused lectures, hands on practical sessions of the theoretical concepts using real life datasets and report writing and project supervision. The domain of applications of Data Analytics is rapidly expanding and there are enormous opportunities for M.Sc. Data Analytics graduates in all these fields.

The duration of the program shall be 4 semesters. The duration of each semester shall be 90 working days. Students may be permitted to complete the program, in a period of 4 continuous semesters from the date of commencement of the first semester of the programs.

Aims and Objectives of the Programme

As Data Analytics is a broad and highly applied multi-disciplinary subject, this syllabus at PG level had set the following aims while preparing the learning and evaluation tools:

1. Introduce Data Analytics as a multi-disciplinary branch of science for solving everyday problems by analyzing relevant data.
2. Introduce a curriculum that imparts the real spirit with which a beginner may approach the learning of any scientific stream, not alone Data Analytics.
3. A curriculum that stresses in equipping the learners with using tools and techniques in Mathematics, Statistics and Computer Science, collaboratively, to solve well defined data analysis problems and present their theoretical work, both in oral and written format to various audiences.
4. A curriculum that stresses the application-oriented approach rather than the traditional theory-based approach.
5. A curriculum which motivates the learners to continue their future study or employment in a very competent manner.
6. A curriculum that attracts the fresher's in Data Analytics to the World of Data Analytics, where numbers are transformed into information.

Syllabus for M.Sc. in Data Analytics

Semester	Course-code	Course Name	Type of the course	Teaching Hours Per Week Theory+ Practical	Credit	Total Credits
I	PG20DA101	Mathematics for Data Analytics	Theory	4+1	3	19
	PG20DA102	Statistics for Data Analytics	Theory	3+2	3	
	PG20DA103	Stochastic Processes & Time Series Analysis	Theory	3+2	3	
	PG20DA104	Introduction to DBMS	Theory	3+2	2	
	PG20DA105	Statistical Programming-I	Theory	2+3	2	
	PG20DA106	Practical – 1 (104,105)	Practical		3	
	PG20DA107	Practical – 2 (101,102,103)	Practical		3	
II	PG20DA201	Applied Regression Analysis	Theory	4+1	3	19
	PG20DA202	Applied Multivariate Analysis	Theory	3+2	3	
	PG20DA203	Machine Learning	Theory	3+2	3	
	PG20DA204	Web Scraping & Text Mining	Theory	3+2	2	
	PG20DA205	Statistical Programming-II	Theory	2+3	2	
	PG20DA206	Practical – 3 (204,205)	Practical		3	
	PG20DA207	Practical – 4 (201,202,203)	Practical		3	
III	PG20DA301	Sampling & Design of Experiments	Theory	4+2	3	21
	PG20DA302	Survival Analysis	Theory	4+2	3	
	PG20DA303	Business Intelligence & Analytics	Theory	5+2	3	
	PG20DA304	Optimization Techniques	Theory	4+2	3	
	PG20DA305	Internship			3	

	PG20DA306	Practical – 5 (303,304)	Practical		3	
	PG20DA306	Practical – 6 (301,302)	Practical			
IV	PG20DA401	Natural Language Processing	Theory	4+2	4	21
	PG20DA402	Elective		3+2	3	
	PG20DA403	Practical – 7 (401,402)			2	
	PG20DA404	Industrial Project		14	10	
	PG20DA405	Project Presentation & Viva Voce			2	
	Total					80

Course-code	List of elective courses
PG20DA402E1	Big Data Analytics Using Hadoop
PG20DA402E2	Artificial Intelligence
PG20DA402E3	Big Data Analytics Using Spark
PG20DA402E4	Fraud Analytics
PG20DA402E5	Neural Networks & Deep Learning
PG20DA402E6	Bayesian Inference
PG20DA402E7	Probabilistic Graphical Models
PG20DA402E8	Complex Networks

Detailed Syllabus

PG20DA101: Mathematics for Data Analytics

Objectives: To make students familiar with various concepts in linear algebra and matrices and their applications in data science, data analytics and pattern recognition.

Outcomes: After completing this course students (i) have understood the properties of Vector spaces

(ii) are able to use the properties of Linear Maps in solving problems in Linear Algebra

(iii) demonstrate proficiency on the topics Eigen values, Eigen vectors and Inner Product Space and can apply linear algebra for applications in Data Analytics.

Module 1

Matrices: Special types of matrices. Systems of linear equations. Elementary operations: Gaussian elimination and row operations, Echelon form of a matrix, Elementary matrices and rank of a matrix, Existence of solution of AX=B. Matrix Algebra: Properties of determinants, Cofactor expansion, Inverse of a matrix, Eigen values and Eigenvectors, eigen values and vectors of special types of matrices, Eigenvectors and Upper Triangular matrices, Quadratic forms, diagonal reduction, canonical forms, nature of definiteness, orthogonal reduction, similarity. Spectral Decomposition of symmetric matrices, Jordan Canonical Form, LU Decomposition, Singular value decomposition

Module 2

Introduction to Vector Spaces: Group, Ring, Field, Vector spaces, Subspaces, sums of Subspaces, Direct Sums, Span and Linear Independence, basis, dimension. Examples, row subspace, column subspace, row rank, column rank, rank of a matrix, inverse of a non-singular matrix, generalized inverse computation and applications, solution of system of linear equations, null space, nullity. Eigen spaces and Diagonal Matrices, Inner Products and Norms, Linear functional on Inner product spaces.

Module 3

Linear transformation and matrices: Definition of Linear Maps- Algebraic Operations on L (V,W)- Null spaces and Injectivity –Range and Surjectivity – Fundamental Theorems of Linear Maps-Representing a Linear Map by a Matrix-Invertible Linear Maps- Isomorphic Vector

spaces-Linear transformation as Matrix Multiplication – Operators – Products of Vector Spaces – Product of Direct Sum – Quotients of Vector spaces

Module 4

Introduction to Tensor analysis: Why Tensor Calculus, Coordinate systems and role of Tensor calculus, change of coordinates, the tensor property-definitions and essential ideas, fundamental properties of tensors, elements of linear algebra in tensor notation.

References

1. Agarwal, R. P., &Flaut, E. C. (2017). *An introduction to linear algebra*. Chapman and Hall/CRC.
2. Anton, H., &Rorres, C. (2013). *Elementary linear algebra: applications version*. John Wiley & Sons.
3. Grinfeld, P. (2013*). Introduction to Tensor Analysis and the Calculus of Moving Surfaces.* Switzerland: Springer New York.
4. Strang, G.(2007). *Linear Algebra and Its Applications.* Cengage learning.
5. Axler, S. J. (1997). *Linear algebra done right* (Vol. 2). New York: Springer.
6. Mathai, A. M. (1997). *Jacobians of matrix transformations and functions of matrix arguments*. World Scientific Publishing Company.

PG20DA102: Statistics for Data Analytics.

Objectives: This course is designed to introduce the concepts of theory of probability, random variables, probability distributions, estimation and testing of hypothesis. This paper also deals with the concept of parametric tests for large and small samples. It provides knowledge about non-parametric tests and its applications. An introduction to the Bayesian concepts in Statistics is also provided. It is also expected to give lab illustration of the concepts through original data sets.

Outcomes (i) Demonstrate the concepts of probability theory, random number generation, distribution theory, sampling distributions, point and interval estimation of unknown parameters and their significance using large and small samples. (ii) Apply the idea of sampling distributions of different statistics in testing of hypotheses. (iii) To understand and apply nonparametric tests for single sample and two samples. (iv) To familiarize the students with Bayesian philosophy.

Module 1

Probability Theory & Random Variables: Basic elements of probability. Introduction to random variables and probability distributions. Univariate distributions- Binomial, Poisson, Geometric, Exponential, Gamma, Beta, Normal and Lognormal distributions, Sampling Distributions and their properties, Random number generation- Basic principles of Random number generation, inversion method, accept-reject method, Random number generation from common distributions. Gibbs sampling, EM algorithm, variational inference

Module 2

Point Estimation: Concepts of Estimation, Estimators and Estimates. Point and interval estimation. Properties of good estimators- unbiasedness, efficiency, consistency and sufficiency.

Methods of Estimation: Methods of moments, maximum likelihood. Invariance property of ML Estimators (without proof). minimum variance. Interval Estimation, 100(1-α)% confidence intervals for mean, variance, proportion, difference of means, proportions and variances.

Module 3

Testing of Hypotheses: Basic ideas of testing of hypotheses, significance level, power, p-value, Neyman-Pearson fundamental Lemma, Distributions with monotone likelihood ratio – Problems, Generalization of the fundamental lemma to randomized tests, uniformly most powerful tests, two sided hypotheses – testing the mean and variance of a normal distribution, testing equality of means and variances of two normal distributions. Likelihood ratio tests –

locally most powerful tests, Large and small sample tests. ANOVA-One way and two way. Sequential Probability Ratio Tests.

Non-Parametric Tests: Introduction to Non parametric tests, Non parametric equivalent of parametric tests.

Module 4

Bayesian statistics: Bayesian parametric models, conjugate prior, Bayesian estimators – Hypothesis testing: testing framework, parametric testing, permutation test, multiple testing.

References

1. Mitzenmacher, M., &Upfal, E. (2017). *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis*. Cambridge university press.
2. Berger, J. O. (2013). *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media.
3. Rohatgi, V. K., & Saleh, A. M. E. (2015). *An introduction to probability and statistics*. John Wiley & Sons.
4. Goon A. M., Gupta M. K., & Dasgupta B.(2005). *Fundamentals of Statistics*, Vol.I, 8th edition, World Press, Kolkatta.
5. Raaiffe H. & Schlaiffer R. (2000) *Applied Statistical Decision Theory*, M.T.Press.

PG20DA103: Stochastic Processes & Time Series Analysis

Objectives: To introduce the basics of stochastic processes and modelling as well as enable the students to analyse time series data and apply suitable techniques to model them and forecast future values.

Outcomes: Students are aware of various stochastic models and time series models and can apply these to model data for predicting future values to make appropriate planning and decision making.

Module 1

Introduction to stochastic processes:- classification of stochastic processes, wide sense and strict sense stationary processes, processes with stationary independent increments, Markov process, Markov chains- transition probability matrices, Chapman-Kolmogorov equation, first passage probabilities, recurrent and transient states, mean recurrence time, stationary distributions, limiting probabilities, Random walk, random walk approximation of Brownian motion and diffusion process, Galton-Watson branching process, generating function relations, mean and variance, extinction probabilities.

Module 2

Continuous time Markov chains, Poisson processes, properties, inter-arrival time distribution pure birth processes and the Yule processes, birth and death processes, Kolmogorov differential equations, linear growth process with immigration, steady-state solutions of Markovian queues - M/M/1, M/M/s, M/MM/∞ models, Renewal processes - concepts, examples, Poisson process viewed as a renewal process.

Module 3

Time series data, examples, Time series as stochastic process, Additive and multiplicative models, stationary time series- covariance stationarity, Modelling Time Series Data, Exponential Smoothing Methods - First-Order Exponential Smoothing, Second and higher Order Exponential Smoothing, Forecasting, Exponential Smoothing for Seasonal Data, Exponential Smoothers.

Module 4

Time series modelling, Autocorrelation function (ACF), partial auto correlation function (PACF), correlogram, AR, MA, ARMA, ARIMA Models, Yule- Walker equations, Box-Jenkins Model fitting and diagnostics. Forecasting future values, Non-linear time series models, ARCH and GARCH Process, GARCH process for modelling volatility, Non-Gaussian Time series modelling, Multivariate Time series analysis.

References

1. Medhi J. (2017) *Stochastic Processes*, Second Edition, Wiley Eastern, New Delhi
2. Ross S.M. (2007) *Stochastic Processes*. Second Edition, Wiley Eastern, New Delhi
3. Feller W. (1968) *Introduction to Probability Theory and its Applications*, Vols. I & II, John Wiley, New York.
4. Karlin S. and Taylor H.M. (1975) *A First Course in Stochastic Processes*, Second edition, Academic Press, New-York.
5. Cinlar E. (1975*) Introduction to Stochastic Processes*, Prentice Hall, New Jersey.
6. Basu A.K. (2003) *Introduction to Stochastic Processes*, Narosa, New-Delhi.
7. Montgomery D. C., Cheryl L. J., and Murat K. (2015) *Introduction to Time Series Analysis and Forecasting*. John Wiley & Sons.
8. Brockwell P.J and Davis R.A. (2002) Introduction to Time Series and Forecasting Second edition, Springer-Verlag.
9. Ruey S. Tsay (2005). *Analysis of Financial Time Series*, Second Ed. Wiley & Sons.
10. Abraham, B., &Ledolter, J. (2009). *Statistical methods for forecasting* (Vol. 234). John Wiley & Sons.
11. Chatfield, C.(2004).*The Analysis of Time Series - An Introduction (Sixth edition*), Chapman and Hall.

PG20DA104: Introduction to DBMS

Objectives: To understand the basic concepts and the applications of database systems.

Outcomes: (i) Students understood the basics of SQL and can construct queries using SQL. (ii) Understood the relational database design principles and the basic issues of transaction processing and concurrency control. (iv) Understood database storage structures and access techniques. (v) Understood object oriented databases, data warehousing and OLAP tools. (vi)Understood MongoDB and can evaluate the nosql databases.

Module 1

Introduction to File and Database systems- History- Advantages, disadvantages- Data views – Database Languages – DBA – Database Architecture – Data Models- Keys – Mapping Cardinalities, Relational Algebra and calculus – Query languages – SQL – Data definition – Queries in SQL – Updates– Views – Integrity and Security – triggers, cursor, functions, procedure – Embedded SQL – overview of QUEL, QBE.

Module 2

Design Phases – Pitfalls in Design – Attribute types –ER diagram – Database Design for Banking Enterprise – Functional Dependence – Normalization (1NF, 2NF, 3NF, BCNF, 4NF, 5NF).File Organization – Organization of Records in files – Indexing and Hashing. Transaction concept – state serializability – Recoverability- Concurrency Control – Locks- Two Phase locking – Deadlock handling – Transaction Management in Multi Databases.

Module 3

Object-Oriented Databases- OODBMS- rules – ORDBMS- Complex Data types – Distributed databases –characteristics, advantages, disadvantages, rules- Homogenous and Heterogeneous Distributed data Storage – XML – Structure of XML Data – XML Document. Introduction to Mongo DB , Overview of NoSQL.

Module 4

Introduction to data warehousing, evolution of decision support systems –Modelling a data warehouse, granularity in the data warehouse – Data warehouse life cycle, building a data warehouse, Data Warehousing Components, Data Warehousing Architecture – On Line Analytical Processing, Categorization of OLAP Tools.

References

1. Silberschatz, A., Korth, H. F., & Sudarshan, S. (1997). *Database System Concepts* (Vol. 4). New York: McGraw-Hill.
2. Pratt, P. J. & Adamski, J. J. (2011). *Database Systems: Management and Design*. Boyd & Fraser Pub. Co.
3. James R Groff and Paul N Weinberg (2003) *The Complete Reference SQL* –, Second Edition, Tata McGraw Hill,
4. Shamkant, R. E., & Navathe, B. (2009). *Fundamentals of Database Systems*. Addison-Wesley Publishing Company.
5. Elmasri, R.& Navathe, S. (2010). *Fundamentals of Database Systems*. Addison-Wesley Publishing Company.)

PG20DA105: Statistical Programming-I

Objectives: To introduce the basics of Statistical programming using the most popular languages R & Python.

Outcomes: (i) Students are able to develop basic programming skills. (ii) Students are able to perform Statistical data analyses using R & Python programming.

Module 1

Introduction to R Programming: Basics of R: Installing the base R system and R-Studio. How to run R, R Sessions and Functions, Basic Math, Variables, Data Types, Vectors, Advanced Data Structures, Data Frames, Lists, Matrices, Arrays, Classes.

Packages in R: Installing and loading packages, familiarizing with popular packages and functions in R, Writing functions in R.

R Programming Structures: Control Statements, Loops, - Looping Over Non-Vector Sets, - If-Else, Arithmetic and Boolean Operators and values, Default Values for Argument, Return Values, Deciding Whether to explicitly call return- Returning Complex Objects.

Module 2

Introduction to Python programming: Designing a program: development cycle, pseudo code, flowcharts and algorithm development; variables, numerical data types and literals, strings, assignment and reassignment, input/output, formatted output, reading numbers and strings from keyboard; performing calculations: floating point and integer division, converting math formulas to programming statements, standard mathematical functions, mixed-type expressions and data type conversions.

Program Decision and Control Structures: Boolean expressions, relational expressions, logical operators, Boolean variables; if, if-else, if-elif-else, inline-if statements, nested structures, and flowcharts; use of temporary variables, application: arranging a few numbers in increasing or non-decreasing, decreasing or non-increasing orders, etc.

Module 3

Repeated calculations and Looping: condition-controlled and count-controlled loops, while loop(condition-controlled), infinite loops; for-loop (count-controlled), applications: calculating summation of series, Taylor expansion of mathematical functions, etc; nested loops.

Arrays, Lists and Tuples: lists, index, iterating over a list with for-loop, operations with lists,

built-in functions, finding index, sorting, etc., processing lists; Arrays: vectors and tuples, vector arithmetic, arrays, Numerical Python arrays – Numpy, curve plotting: matplotlib, SciTools, making animations and videos; Higher-dimensional arrays: two- and three-dimensional arrays, matrix objects and matrix operations: inverse, determinant, solving linear systems using standard libraries.

Module 4

Intermediate Python: Matlibplot, Seaborn, Dictionaries and Pandas, Logic, Control flow and filtering. Importing data in Python, Importing from flat files such as .txts and .csvs, from files native to other software such as Excel spreadsheets, Stata, SAS and MATLAB files, from relational databases such as SQLite & PostgreSQL, from the web and from Application Programming Interfaces, also known as APIs. Cleaning data in Python: Exploring data, tidying data for analysis, combining data for analysis, cleaning data for analysis, case studies. Manipulating data frames with pandas-Extracting and transforming data, advanced indexing, rearranging and reshaping data, grouping data. Graphical exploratory analysis, numerical exploratory analysis.

References

1. VanderPlas, J. (2016). *Python data science handbook: Essential tools for working with data*. " O'Reilly Media, Inc.".
2. Gaddis, T., & Agarwal, R. (2015). *Starting out with Python*. Pearson.
3. Langtangen, H. P. (2014). *A primer on scientific programming with Python* (Vol. 6). Springer.
4. Grus, J. (2019). *Data science from scratch: first principles with python*. O'Reilly Media.

PG20DA201: Applied Regression Analysis

Objectives: To introduce the various concepts and techniques in regression analysis for modelling data and forecasting future values.

Outcomes: The students have studied simple linear regression, multiple regression, and residual analysis for fitting a suitable model to a given data and to check the suitability. They have studied necessary transformations and modifications to be made when model assumptions are violated. They are capable of fitting logistic and Poisson models, orthogonal and polynomial models. They have understood ridge regression, kernel regression, non-parametric regression etc.

Module 1

Introduction to regression analysis: overview and applications of regression analysis, major steps in regression analysis. Simple linear regression (Two variables): assumptions, estimation and properties of regression coefficients, significance and confidence intervals of regression coefficients, measuring the quality of the fit. Residual analysis, various types of residuals, Departures from underlying assumptions, Departures from normality, Diagnostics and remedies,

Module 2

Multiple linear regression model: assumptions, ordinary least square estimation of regression coefficients, interpretation and properties of regression coefficient, significance and confidence intervals of regression coefficients. Mean Square error criteria, coefficient of determination, criteria for model selection; Need for transformation of variables; power transformation, Box-Cox transformation; removal of heteroscedasticity and serial correlation, Leverage and influence. Effect of outliers.

Module 3

Generalized least squares and weighted least squares. Polynomial regression models, Forward, Backward and Stepwise procedures. Nonparametric regression, Kernel regression, Loess, ridge regression, orthogonal polynomials, indicator variables**,** subset regression, stepwise regression, variable selection Robust regression.

Module 4

Introduction to nonlinear regression, linearity transformations, logarithmic transformation, Least squares in the nonlinear case and estimation of parameters, Models for binary response variables, generalized linear models, estimation and diagnosis methods for Logistic and

Poisson regressions. Prediction and residual analysis, Multinomial logistic regression, Random and mixed effect models, Multi-collinearity, sources, effects, tests.

References

1. D. C Montgomery, E.A Peck and G.G Vining (2003). *Introduction to Linear Regression Analysis*, John Wiley and Sons, Inc.NY,
2. S. Chatterjee and A. Hadi (2013) *Regression Analysis by Example*, 5th Ed., John Wiley and Sons.
3. Seber, A.F. and Lee, A.J. (2003) *Linear Regression Analysis*, John Wiley, Relevant sections from
4. Iain Pardoe (2012) *Applied Regression Modelling*, John Wiley and Sons, Inc,.
5. P. McCullagh, J.A. Nelder,(1989) *Generalized Linear Models*, Chapman & Hall,. John O. Rawlings,
6. Sastry G. Pantula, David A. Dickey (1998) *Applied Regression Analysis*, Second Edition, Springer.
7. Draper, N. and Smith, H. (2012) *Applied Regression Analysis* – John Wiley & Sons

PG20DA202: Applied Multivariate Analysis

Objectives: To introduce multivariate data and associated concepts, distributions, testing and estimation, and the theory and applications of discriminant function and classification rules, principal components, canonical correlations, factor analysis etc.

Outcomes: After undergoing this course students can apply multivariate techniques such as discriminant function and classification rules, principal components, canonical correlations, factor analysis, MANOVA etc. They are enabled to apply Hotelling‘s T2 and Mahalanobis D2 etc for testing hypotheses in the case of multivariate data.

Module 1

Basic concepts on multivariate variable. Concept of random vector: Its expectation and dispersion (Variance-Covariance) matrix. Marginal and joint distributions. Conditional distributions and Independence of random vectors. Multinomial distribution. Multivariate normal distribution, Marginal and conditional distribution, characteristic function, additive property, mle.s of mean and dispersion matrix.

Module 2

Sample mean vector and its distribution, Hotelling‘s T2 and Mahalanobis‘ D2 statistics and applications. Tests of hypotheses about the mean vectors and covariance matrices for multivariate normal populations. Wishart distribution, Rao‘s U, Pillai‘s Trace statistics; Independence of sub vectors and sphericity test.

Module 3

Bayes minimax and Fisher‘s criteria for discrimination between two multivariate normal populations. Sample discriminant function. Tests associated with discriminant functions. Probabilities of misclassification and their estimation. Discrimination for several multivariate normal populations. Multivariate analysis of variance (MANOVA) of one and two- way classified data. Multivariate analysis of covariance, illustrative numerical examples.

Module 4

Principal components, sample principal components asymptotic properties. Canonical variables and canonical correlations: definition, estimation, computations. Factor analysis: Orthogonal factor model, factor loadings, estimation of factor loadings, factor scores, cluster analysis-agglomerative and divisive techniques. Applications to real data sets and problems.

References

1. Chatfield, C. (2018). *Introduction to multivariate analysis*. Routledge.
2. Rencher, A. C. (2012) *Methods of Multivariate Analysis*.(3rd ed.) John Wiley.
3. Johnson R.A. and Wichern D.W. (2008) *Applied Multivariate Statistical Analysis*. 6th Edition, Pearson Education.
4. Anderson, T.W. (2009). *An Introduction to Multivariate Statistical Analysis*, 3rd Edition, John Wiley.
5. Everitt B, Hothorn T,(2011). *An Introduction to Applied Multivariate Analysis with R*, Springer.
6. Barry J. Babin, Hair, Rolph E Anderson, & William C. Blac, (2013), *Multivariate Data Analysis*, Pearson New International Edition,

PG20DA203: Machine Learning

Objectives: To familiarize the students with various aspects of machine learning techniques and their applications.

Outcomes: The students have understood different techniques such as unsupervised learning, dimensionality reduction, PCA, SVM, Discriminant function, multilayer preceptors, cluster analysis etc

Module 1

Machine Learning-Examples of Machine Applications – Learning Associations – Classification- Regression- Unsupervised Learning- Reinforcement Learning. Supervised Learning: Learning class from examples- Probably Approximately Correct (PAC) Learning- Noise-Learning Multiple classes. Regression – Model Selection and Generalization.

Introduction to Parametric methods- Maximum Likelihood Estimation: Bernoulli Density-Multinomial Density-Gaussian Density, Nonparametric Density Estimation: Histogram Estimator-Kernel Estimator-K-Nearest Neighbour Estimator.

Module 2

Dimensionality Reduction: Introduction- Subset Selection- Principal Component Analysis, Feature Embedding-Factor Analysis- Singular Value Decomposition- Multidimensional Scaling-Linear Discriminant Analysis- Bayesian Decision Theory. Linear Discrimination: Introduction-Generalizing the Linear Model- Geometry of the Linear Discriminant- Pairwise Separation-Gradient Descent-Logistic Discrimination. Optical separating hyper plane – v-SVM, kernel tricks – vertical kernel- vertical kernel- defining kernel- multiclass kernel machines- one-class kernel machines.

Module 3

Multilayer perceptron, Introduction, training a perceptron- learning Boolean functions-multilayer perceptron- back propagation algorithm- training procedures. Combining Multiple Learners, Rationale-Generating diverse learners- Model combination schemes- voting, Bagging- Boosting- fine tuning an Ensemble.

Module 4

Cluster Analysis, Introduction-Mixture Densities, K-Means Clustering- Expectation-Maximization algorithm- Mixtures of Latent Variable Models-Supervised Learning after Clustering-Spectral Clustering- Hierarchical Clustering- Divisive Clustering- Choosing the number of Clusters.

References

2. James, G., Witten, D., Hastie, T., &Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, pp. 3-7). New York: springer.
3. James, G., Witten, D., Hastie, T., &Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, pp. 3-7). New York: springer.
4. Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.
5. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
6. Ian, G., Yoshua, B., & Aaron, C.(2006).*Deep Learning*, MIT Press.
7. Ethem,A.(2004).*Introduction to Machine Learning (Adaptive Computation and Machine Learning Series).* The MIT Press.
8. Alpaydin (2014) *Introduction to Machine Learning*, 3rd Edition, MIT Press.
9. Frank Kane (2012) *Data Science and Machine Learning*. Manning Publications.
10. C.M.Bishop, Pattern Recognition and Machine Learning, Springer.
11. T. Hastie, R. Tibshirani and J. Friedman (2016) *The Elements of Statistical Learning*: *Data Mining, Inference and Prediction*, Springer, 2nd Edition,2009

PG20DA204: Web Scraping & Text Mining

Objectives: To introduce web and data technologies like HTML, XML, JSON, Xpath, AJAX etc for web scrapping and text mining.

Outcomes: After this course students are well equipped with tools for web scrapping and text mining and can apply these in real contexts.

Module 1

Introduction to Web and Data Technologies: HTML, Browser presentation and source code, Syntax rules, Tags and attributes, Parsing,; XML and JSON : A short example XML document, XML syntax rules, When is an XML document well formed or valid?, XML extensions and technologies, XML and R in practice, A short example JSON document, JSON syntax rules, JSON and R in practice,

Module 2

Xpath: Xpath—a query language for web documents, Identifying node sets with Xpath, Extracting node elements, HTTP: HTTP fundamentals, Advanced features of HTTP, Protocols beyond HTTP, HTTP in action,

Module 3

AJAX: JavaScript, XHR, Exploring AJAX with Web Developer Tools, SQL and Relational Databases: Overview and terminology, Relational Databases, SQL: a language to communicate with Databases, R packages to manage databases

Module 4

Web Scraping and Text Mining: Scraping the Web**:** Retrieval scenarios, Scraping data from AJAX-enriched webpages with Selenium/Rwebdriver, Retrieving data from APIs, Authentication with Oauth, Extraction strategies, Web scraping: Good practice, Statistical Text Processing: The running example: Classifying press releases of the British government, Processing textual data, Supervised learning techniques, Unsupervised learning techniques, Managing Data Projects: Interacting with the file system, Organizing scraping procedures, Executing R scripts on a regular basis.

References

1. Munzert, S.,Rubba, C.,MeiBner, P., Nyhuis, D. (2014) *Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining*. John Wiley & Sons,

PG20DA205: Statistical Programming-II

Objectives: To introduce advanced concepts of Statistical programming using the most popular programming language Python.

Outcomes: (i) Students are able to develop advanced programming skills. (ii) Students are able to perform advanced Statistical data analyses using Python programming.

Module 1

Linear Regression – Introduction – Regression model for one variable regression – Selecting best model – Error measures SSE, SST, RMSE, R2 – Interpreting R2 – Multiple linear regression – Loess and ridge regression – Correlation – Recitation – A minimum of 3 data sets for practice.

Logistic Regression – The Logit – Confusion matrix – sensitivity, specificity – ROC curve – Threshold selection with ROC curve – Making predictions – Area under the ROC curve (AUC) - Recitation – A minimum of 3 data sets for practice

Module 2

Approaches to missing data – Data imputation – Multiple imputation – Classification and Regression Tress (CART) – CART with Cross Validation – Predictions from CART – ROC curve for CART – Random Forests – Building many trees – Parameter selection – K-fold Cross Validation – Recitation – A minimum of 3 data sets for practice

Module 3

Time series analysis – Clustering – k-mean clustering – Random forest with clustering – Understanding cluster patterns – Impact of clustering – Heatmaps –Factor Analysis- Discriminant Analysis-Principal Component Analysis- Recitation – min 3 data sets for practice

Module 4

Support Vector Machines – Gradient Boosting – Naïve Bayes – Bayesian GLM – GLMNET – Ensemble modelling – Experimenting with all of the above approaches with and without data imputation and assessing predictive accuracy – Recitation – minimum 3 data sets for practice.

References

1. VanderPlas, J. (2016). *Python data science handbook: Essential tools for working with data*. " O'Reilly Media, Inc.".
2. Gaddis, T., & Agarwal, R. (2015). *Starting out with Python*. Pearson.

3. Langtangen, H. P. (2014). *A primer on scientific programming with Python* (Vol. 6). Springer.
4. Grus, J. (2019). *Data science from scratch: first principles with python*. O'Reilly Media.

PG20DA301: Sampling & Design of Experiments

Objectives: To make the students familiar with different sampling schemes and their advantages as well as their applications in estimating population mean, total, proportion etc. It is also expected to impart knowledge on basic principles of experimentation, different designs like CRD, RBD, LSD, BIBD, Factorial experiments etc.

Outcomes: (i) After undergoing this course, students are aware of different sample survey methods and are capable of planning and implementing sample surveys, consumer satisfaction surveys, public opinion surveys etc.(ii) they are aware of different designs in experimentation like CRD, RBD, LSD, BIBD, Factorial Designs, etc. and can apply ANOVA technique to analyse the data using Python or R.

Module 1

Sampling 1: Census and sampling methods, probability sampling and non-probability sampling, principal steps in sample surveys, sampling errors and non-sampling errors, bias, variance and mean square error of an estimator, simple random sampling with and without replacement, estimation of the population mean, total and proportions, properties of the estimators, variance and standard error of the estimators, confidence intervals, determination of the sample size. Stratified random sampling, estimation of the population mean, total and proportion, properties of estimators, various methods of allocation of a sample, comparison of the precisions of estimators under proportional allocation, optimum allocation and SRS. Systematic sampling – Linear and Circular, estimation of the mean and its variance. Comparison of systematic sampling, SRS and stratified random sampling for a population with a linear trend.

Module 2

Sampling 2: Ratio method of estimation, estimation of population ratio, mean and total, Bias and relative bias of ratio estimator, comparison with SRS estimation. Regression method of estimation. Comparison of ratio and regression estimators with mean per unit method, Cluster sampling, single stage cluster sampling with equal and unequal cluster sizes, estimation of the population mean and its standard error. Multistage and Multiphase sampling (Basic Concepts), estimation of the population mean and its standard error, Reservoir sampling (Basic concepts)

Module 3

Linear estimation: standard Gauss Markoff set up, estimability of parameters, method of least squares, best linear unbiased Estimators, Gauss – Mark off Theorem, tests of linear hypotheses.

Planning of experiments, Basic principles of experimental design, uniformity trails, analysis of variance, one-way, two-way and three-way classification models, completely randomized design (CRD), randomized block design (RBD) Latin square design (LSD) and Graeco-Latin square designs, Analysis of covariance (ANCOVA), ANCOVA with one concomitant variable in CRD and RBD.

Module 4

Incomplete block design: balanced incomplete block design (BIBD); incidence Matrix, parametric relation; intrablock analysis of BIBD, basic ideas of partially balanced incomplete block design (PBIBD). Factorial experiments, 2n and 3n factorial experiments, analysis of 22, 23 and 32 factorial experiments, Yates procedure, confounding in factorial experiments, basic ideas of response surface designs.

References

1. Cochran W. G. (1999) *Sampling Techniques*, 3rd edition, John Wiley and Sons.
2. Mukhopadyay P. (2009) *Theory and Methods of Survey Sampling*, 2nd edition, PHL, New Delhi.
3. Aloke Dey (1986) *Theory of Block Designs*, Wiley Eastern, New Delhi.
4. Das M.N. and Giri N.C. (1994) *Design and analysis of experiments*, Wiley Eastern Ltd.
5. Arnab, R. (2017). *Survey Sampling: Theory and Applications*. Academic Press.
6. Montgomery, C.D. (2012) *Design and Analysis of Experiments*, John Wiley, New York.
7. Sampath S. C. (2001*) Sampling Theory and Methods*, Alpha Science International Ltd.,
8. Thomas Lumley (1996) *Complex Surveys. A Guide to Analysis Using R*, Wiley eastern Ltd.
9. Des Raj (1967*) Sampling Theory*. Tata McGraw Hill ,NewDelhi
10. Dean, A. and Voss, D. (1999) *Design and Analysis of Experiments*, Springer Texts in Statistics

PG20DA302: Survival Analysis

Objectives: Survival Analysis is highly applied in clinical data. This course will help the students in handling clinical data and related analysis.

Outcomes: The students are able to perform Statistical analyses on clinical data.

Module 1

Basic Quantities and Models - Survival function, Hazard function, Mean residual life function and Median life, Common Parametric Models for Survival Data; Censoring and Truncation - Right Censoring, Left or Interval Censoring, Truncation, Likelihood Construction for Censored and Truncated Data

Module 2

Nonparametric Estimation of a Survivor Function and Quantiles, The Product-Limit Estimator, Nelson-Aalen Estimator, Interval Estimation of Survival Probabilities or Quantiles, Asymptotic Properties of Estimators, Descriptive and Diagnostic Plots, Plots Involving Survivor or Cumulative Hazard Functions, Classic Probability Plots, Estimation of Hazard or Density Functions, Methods for Truncated and Interval Censored Data, Left-Truncated Data, Right-Truncated Data, Interval-Censored Data.

Module 3

Semi-parametric Proportional Hazards Regression with Fixed Covariates – Coding Covariates, Partial Likelihoods for Distinct-Event Time Data, Partial Likelihoods when Ties are present, Local Tests, Discretizing a Continuous Covariate, Model Building using the Proportional Hazards Model, Estimation for the Survival Function; Introduction to Time-Dependent Covariates; Regression Diagnostics :- Cox-Snell Residuals for assessing the fit of a Cox Model, Graphical Checks of the Proportional Hazards Assumption, Deviance Residuals, Checking the Influence of Individual Observations

Module 4

Inference for Parametric Regression Models - Exponential, Gamma and Weibull Distributions, Nonparametric procedure for comparison of survival function, Competing risk models – Basic Characteristics and Model Specification.

References

1. Klein J.P. and Moeschberger M.L. (2003*) Survival Analysis - Techniques for censored and truncated data,* Second Edition, Springer-Verlag , New York,
2. Lawless J.F (2003*) Statistical Models and Methods for Lifetime Data*, Second Editon, John Wiley & Sons
3. Kalbfleisch J.D and Prentice, R.L. (2002*) The Statistical Analysis of Failure Time Data*, Second Edition, John Wiley & Sons Inc.
4. Hosmer Jr. D.W and Lemeshow S (1999) *Applied Survival Analysis – Regression Modelling of Time to event Data*, John Wiley & Sons. Inc.
5. Nelson. W (1982) *Applied Life Data Analysis*.
6. Miller, R.G. (1981) *Survival Analysis*, John Wiley.

PG20DA303: Business Intelligence & Analytics

Objectives: to introduce various concepts in business intelligence, data warehousing, data mining, business models, knowledge management, data extraction, data life cycle.

Outcomes: Understood the tools for business intelligence and total quality management. Students are now familiar with basic methodology in data warehousing and data mining as well as knowledge management. Now they are capable of applying these for data extraction, strategy development, business intelligence etc.

Module 1:

Business Intelligence: Introduction, Definition, Business Intelligence Segments, Difference between Information and Intelligence, Defining Business Intelligence Value Chain, Factors of Business Intelligence System, Real time Business Intelligence, Business Intelligence Applications.

Creating Business Intelligence Environment, Business Intelligence Landscape, Types of Business Intelligence, Business Intelligence Platform, Dynamic roles in Business Intelligence, Roles of Business Intelligence in Modern Business- Challenges of BI. Business Intelligence Types: Introduction, Multiplicity of Business Intelligence Tools, Types of Business Intelligence Tools, Modern Business Intelligence, the Enterprise Business Intelligence, Information Workers. Architecting the Data: Types of Data, Enterprise Data Model, Enterprise Subject Area Model, Enterprise Conceptual Model,

Enterprise Conceptual Entity Model, Granularity of the Data, Data Reporting and Query Tools, Data Partitioning, Meta data, Total Data Quality Management (TDQM).

Module 2:

Introduction to Data Mining: Definition of Data Mining, Architecture of Data Mining, Kinds of Data which can be mined, Functionalities of Data Mining, Classification on Data Mining system, Various risks in Data Mining, Advantages and disadvantages of Data Mining, Ethical issues in Data Mining, Analysis of Ethical issues.

Introduction to Data Warehousing: Introduction, Advantages and Disadvantages of Data Warehousing, Data Warehouse, Data Mart, Aspects of Data Mart, Online Analytical Processing, Characteristics of OLAP, OLAP Tools, OLAP Data Modelling, OLAP Tools and the Internet, Difference between OLAP and OLTP, Multidimensional Data Model

Module 3:

Types of Business Models, B2B Business Intelligence Model, Electronic Data Interchange & E-Commerce Models, Advantages of E-Commerce for B2B Businesses, Systems for Improving

B2B E-Commerce, B2C Business Intelligence Model, Need of B2C model in Data warehousing, Different types of B2B intelligence Models.

Knowledge Management: Introduction, Characteristics of Knowledge Management, Knowledge assets, Generic Knowledge Management Process, Knowledge Management Technologies, Essentials of Knowledge Management Process

Module 4:

Data Extraction: Introduction, Data Extraction, Role of ETL process, Importance of source identification, Various data extraction techniques, Logical extraction methods, Physical extraction methods, Change data capture

Business Intelligence Life Cycle: Introduction, Business Intelligence Lifecycle, Enterprise Performance Life Cycle (EPLC)Framework Elements, Life Cycle Phases, BI Strategy, Objectives and Deliverables, Transformation Roadmap, Building a transformation roadmap, BI Development Stages and Steps, Parallel Development Tracks, BI Framework.

References

1. Tan, P. N., Steinbach, M., & Kumar, V. (2016). *Introduction to data mining*. Pearson Education India.
2. Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. San Francisco, CA, itd.
3. Soman, K. P., Diwakar, S., & Ajay, V. (2006). *Data mining: theory and practice [with CD]*. PHI Learning Pvt. Ltd.
4. Alex, B., & Stephen, J. S..(2004).*Data Warehousing*, *Data Mining & OLAP*.Mcgraw-Hill, Tata McGraw-Hill Education.
5. Business Intelligence Guidebook: From Data Integration to Analytics by Rick Sherman
6. Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications by Larissa T. Moss and Shaku Atre
7. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling by Ralph Kimball and Margy Ross
8. Successful Business Intelligence, Second Edition: Unlock the Value of BI & Big Data by Cindi Howson
9. Business Intelligence for Dummies by Swain Scheps

PG20DA304: Optimization Techniques

Objectives: To make the students familiar with modern optimization techniques.

Outcomes: (i) After undergoing this course, students are aware of different modern approaches in numerical optimization methods.

Module 1

Introduction to optimization: formulation of optimization problems-Review of classical methods- Linear programming- Nonlinear programming- Constraint optimality criteria-constrained optimization- Population based optimization techniques

Module 2

Genetic Algorithm-Introduction: Working principle- Representation-selection-fitness assignment reproduction-cross over-mutation-constraint handling-advanced genetic algorithms-Applications- Artificial Immune Algorithm-Introduction-Clonal selection algorithm-Negative selection algorithm-Immune network algorithms-Dendritic cell algorithms

Module 3

Differential Evolution: Introduction-Working principles-parameter selection-advanced algorithms in Differential evolution- Biogeography- Based Optimization- Introduction-Working principles - Algorithmic variations.

Module 4

Particle Swarm Optimization: Introduction- Working principles- Parameter selection-Neighbourhoods and Topologies-Convergence-Artificial Bee Colony Algorithm-Introduction-Working principles- Applications Cuckoo search based algorithm-Introduction- Working principles- Random walks and the step size-Modified cuckoo search

Hybrid Algorithms: Concepts- divide and conquer- decrease and conquer-HPABC-HBABC-HDABCHGABC-Shuffled Frog Leaping Algorithm-- Working principles -Parameters- Grenade Explosion Algorithm-Working principle-Applications.

References

1. Rao, S. S. (2019*). Engineering optimization: theory and practice.* John Wiley & Sons.
2. Venkata Rao, R. (2016). *Teaching Learning Based Optimization Algorithm*: And Its Engineering Applications, 1e, Springer.
3. Simon, D. (2013). *Evolutionary optimization algorithms*. John Wiley & Sons.
4. Yang, X. S. (2010). *Engineering optimization: an introduction with metaheuristic applications*. John Wiley & Sons.

PG20DA401: Natural Language Processing

Objectives: To introduce various techniques for natural language processing using Python.

Outcomes: Now students are aware of Text classification, Text summarization, Semantic and sentiment analysis, classification algorithms and can apply these in practical real life problems.

Module 1

What is Natural Language Processing? Language Processing and Python, Natural Language Basics, Natural Language, Linguistics, Language Syntax and Structure, Language Semantics, Text Corpora, Natural Language Processing, Text Analytics; Processing and Understanding Text: Text Tokenization, Text Normalization, Understanding Text Syntax and Structure

Module 2

Text Classification: What Is Text Classification? Automated Text Classification, Text Classification Blueprint, Text Normalization, Feature Extraction, Classification Algorithms, Evaluating Classification Models, Building a Multi-Class Classification System

Module 3

Text Summarization: Text Summarization and Information Extraction, important concepts, Text Normalization Feature Extraction, Key phrase Extraction, Topic Modelling, Automated Document Summarization; Text Similarity and Clustering: Important Concepts, Text Normalization, Feature Extraction, Text Similarity, Analysing Term Similarity, Analysing Document Similarity, Document Clustering

Module 4

Semantic and Sentiment Analysis: Semantic Analysis, Exploring WordNet, Word Sense Disambiguation, Named Entity Recognition, Analysing Semantic Representations; Stemming and Lemmatization, Synsets and Hypernyms.

References

1. Dipanjan Sarkar, *Text Analytics with Python*, Apress/Springer, 2016
2. Jurafsky, D., & Martin, J. H. (2014). *Speech and language processing*. Vol. 3. Prentice Hall.
3. Bird, S., Klein, E., &Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc.".
4. Manning, C. D., Manning, C. D., &Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT press.

ELECTIVE

PG20DA402E1: Big Data Analytics Using Hadoop

Objectives: To make the students aware about the different techniques for big data analytics.

Outcomes: After undergoing this course students are enabled to use Hadoop, RDBMS, Mapreduce, HDFS, HIVE & PIG etc for big data analytics.

Module 1

Distributed file system – Big Data and its importance, Four Vs, Drivers for Big data, Big data analytics, Big data applications, Algorithms using map reduce, Matrix-Vector Multiplication by Map Reduce. Apache Hadoop– Moving Data in and out of Hadoop – Understanding inputs and outputs of MapReduce – Data Serialization, Problems with traditional large-scale systems Requirements for a new approach- Hadoop – Scaling-Distributed Framework- Hadoop v/s RDBMS-Brief history of Hadoop. Installing and Configuring Hadoop; Configurations of Hadoop: Hadoop Processes (NN, SNN, JT, DN, TT)-Temporary directory – UI-Common errors when running Hadoop cluster, solutions. Setting up Hadoop on a local Ubuntu host: Prerequisites, downloading Hadoop, setting up SSH, configuring the pseudo-distributed mode, HDFS directory, Name Node, Examples of MapReduce, Using Elastic MapReduce, Comparison of local versus EMR Hadoop.

Module 2

Understanding MapReduce: Key/value pairs, Hadoop Java API for MapReduce, Writing MapReduce programs, Hadoop-specific data types, Input/output. Developing MapReduce Programs: Using languages other than Java with Hadoop, Analyzing a large dataset. Advanced MapReduce Techniques: Simple, advanced, and in-between Joins, Graph algorithms, using language-independent data structures. Hadoop configuration properties – Setting up a cluster, Cluster access control, managing the Name Node, Managing HDFS, MapReduce management, Scaling.

Module 3

Hadoop Streaming - Streaming Command Options – Specifying a Java Class as the Mapper/Reducer– Packaging Files With Job Submissions – Specifying Other Plug-ins for Jobs. HIVE & PIG ; Architecture, Installation, Configuration, Hive vs RDBMS, Tables, DDL & DML, Partitioning & Bucketing, Hive Web Interface, Pig, Use case of Pig, Pig Components, Data Model, Pig Latin.

Module 4

Hbase RDBMS VsNoSQL, Hbasics, Installation, Building an online query application – Schema design, Loading Data, Online Queries, Successful service. Hands On: Single Node Hadoop Cluster Set up in any cloud service provider- How to create instance. How to connect that Instance Usingputty.I nstalling Hadoop framework on this instance. Run sample programs which come with Hadoop
framework.

References:

1. Boris lublinsky, Kevin t. Smith, Alexey Yakubovich, Professional Hadoop Solutions, Wiley, 2015.
2. Tom White, Hadoop: The Definitive Guide, O'Reilly Media Inc., 2015.
3. Garry Turkington, Hadoop Beginner's Guide, Packt Publishing, 2013.
4. Pethuru Raj, Anupama Raman, DhivyaNagaraj and Siddhartha Duggirala, High-Performance Big-
Data Analytics: Computing Systems and Approaches, Springer, 2015.
5. Jonathan R. Owens, Jon Lentz and Brian Femiano, Hadoop Real-World Solutions Cookbook, Packt Publishing, 2013.

PG20DA402E2: Artificial Intelligence

Objectives: To understand thinking and intelligence in ways that enable the construction of computer systems that are able to reason in uncertain environments.

Outcomes: able to articulate and exemplify the basic knowledge artificial intelligence, Understand the basics of knowledge representation, can use AI programming languages and the methods of AI implementation and can recommend AI strategies based on applications

Module 1

Artificial Intelligence - Introduction, AI Problems, AI Techniques, The Level of the Model, Criteria For Success. Defining the Problem as a State Space Search, Problem Characteristics, Production Systems, Search: Issues in The Design of Search Programs, Un-Informed Search, BFS, DFS; Heuristic Search Techniques: Generate-And-Test, Hill Climbing, Best-First Search, A*Algorithm, Problem Reduction, AO*Algorithm, Constraint Satisfaction, Means-Ends Analysis.

Knowledge Representation: Procedural Vs Declarative Knowledge, Representations & Approaches to Knowledge Representation, Forward Vs Backward Reasoning, Matching Techniques, Partial Match - ing, Fuzzy Matching Algorithms and RETE Matching Algorithms;

Module 2

Logic Based Programming-AI Programming languages: Overview of LISP, Search Strategies in LISP, Pattern matching in LISP , An Expert system Shell in LISP, Over view of Prolog, Production System using Prolog; Symbolic Logic: Propositional Logic, First Order Predicate Logic: Representing Instance and Relationships, Computable Functions and Predicates, Syntax & Semantics of FOPL, Normal Forms, Unification &Resolution, Representation Using Rules, Natural Deduction; Structured Representations of Knowledge: Semantic Nets, Partitioned Semantic Nets, Frames, Conceptual Dependency, Conceptual Graphs, Scripts, CYC;.

Module 3

Reasoning under Uncertainty: Introduction to Non-Monotonic Reasoning, Truth Maintenance Systems, Logics for non-monotonic Reasoning, Model and Temporal Logics; Statistical Reasoning: Bayes' Theorem, Certainty Factors and Rule-Based Systems, Bayesian Probabilistic Inference, Bayesian Networks, Dempster -Shafer Theory, Fuzzy Logic: Crisp Sets ,Fuzzy Sets, Fuzzy Logic Control, Fuzzy Inferences & Fuzzy Systems.

Module 4

Experts Systems: Overview of an Expert System, Structure of an Expert Systems, Different Types of Expert Systems-Rule Based, Model Based, Case Based and Hybrid Expert Systems,

Knowledge Ac - quisition and Validation Techniques, Black Board Architecture, Knowledge Building System Tools, Expert System Shells, Fuzzy Expert systems.

1. George F Luger (2016) *Artificial Intelligence*, Pearson Education Publications
2. Elaine Rich and Knight (2017*) Artificial Intelligence*, Mcgraw-Hill Publications
3. Patterson, D.W.(2005) *Introduction to Artificial Intelligence & Expert Systems*, PHI
4. Weiss.G, (2000) *Multi Agent Systems- A Modern Approach to Distributed Artificial Intelligence*, MIT Press.
5. Russell S. and Norvig, P.(2010) *Artificial Intelligence : A modern Approach*, Printice Hall
6. Elaine, R., & Kevin, K. (2017*). Artificial Intelligence, 3e.* Tata McGraw Hill.
7. Russell, S. J., &Norvig, P. (2016). *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited.
8. Khemani, D. (2013). *A first course in artificial intelligence*. McGraw-Hill Education.
9. Michalewicz, Z., & Fogel, D. B. (2013). *How to solve it: modern heuristics*. Springer Science & Business Media.
10. Edelkamp, S., &Schroedl, S. (2011). *Heuristic search: theory and applications*. Elsevier.
11. Winston, P. H.(2002). *Artificial Intelligence, 1e*, Pearson

PD20DA402E3: Big Data Analytics Using Spark

Objectives: To make the students aware about the different techniques for big data analytics.

Outcomes: After undergoing this course students are enabled to use Spark for big data analytics.

Module 1

Introduction to Data Analysis with Spark: Distributed file system – Big Data and its importance, Four Vs, Drivers for Big data, Big data analytics, Big data applications, Introduction to Spark's Python and Scala Shells, Introduction to Core Spark Concepts, Initializing a Spark Contex, Building Standalone, Programming with RDDs, RDD Basics, Creating RDDs, RDD Operations, Transformations, Actions, Lazy Evaluation, Passing Functions to Spark, Common Transformations and Actions, Basic RDDs, Converting Between RDD Types, Persistence (Caching)

Module 2

Working with Key/Value Pairs, Creating Pair RDDs, Transformations on Pair RDDs, Aggregations Grouping Data, Joins, Sorting Data, Actions Available on Pair RDDs, Data Partitioning (Advanced), Determining an RDD's Partitioner, Operations That Benefit from Partitioning, Operations That Affect Partitioning, Custom Partitioners

Module 3

Loading and Saving Data: File Formats-Text Files, JSON, Comma-Separated Values and Tab-Separated Values, Sequence Files, Object Files, Hadoop Input and Output Formats, File Compression, Filesystems- Local/"Regular" FS, Amazon S3, HDFS, Structured Data with Spark SQL, Apache Hive, JSON, Databases- Java Database Connectivity, Cassandra, HBase, Elasticsearch.

Tuning and Debugging Spark: Configuring Spark with SparkConf, Components of Execution: Jobs, Tasks, and Stages, Finding Information, Spark Web UI, Driver and Executor Logs, Key Performance Considerations, Level of Parallelism, Serialization Format, Memory Management, Hardware Provisioning

Module 4

Advanced Spark Programming: Accumulators, Accumulators and Fault Tolerance, Custom Accumulators, Broadcast Variables, Optimizing Broadcasts, Working on a Per-Partition Basis, Piping to External Programs, Numeric RDD Operations,

Machine Learning with MLlib: Overview , Data Types- Working with Vectors, Algorithms- Feature Extraction, Statistics, Classification and Regression, Clustering, Collaborative Filtering and Recommendation, Dimensionality Reduction, Model Evaluation, Preparing Features, Configuring Algorithms , Caching RDDs to Reuse, Recognizing Sparsity, Level of Parallelism, Pipeline API

References

1. Karau, H., Konwinski, A., Wendell, P., &Zaharia, M. (2015). *Learning spark: lightning-fast big data analysis*. " O'Reilly Media, Inc.".
2. Chambers, B., &Zaharia, M. (2018). *Spark: The definitive guide: Big data processing made simple*. " O'Reilly Media, Inc.".
3. Guller, M. (2015). *Big data analytics with Spark: A practitioner's guide to using Spark for large scale data analysis*. Apress.

PG20DA402E4: Fraud Analytics

Objectives: To introduce how one can treat the Internet as a source of data and analyse web-scale data using distributed computing.

Outcomes: Students have gained the basic knowledge on analysis of fraud and fraud detection models; compare different models, develop automation process of fraud detention; and are equipped to formulate and evaluate fraud detection

Module 1

Formulation and evaluation of fraud detection - Fraud detection using data analysis Obtain and cleanse the data for fraud detection-preprocess data for fraud detection - sampling, missing values, outliers, categorization etc.- Explain characteristics and components of the data and assess its completeness

Module 2

Identify known fraud symptoms -and use digital analysis to identify unknown fraud symptoms-Fraud detection models using supervised analytics (logistic regression, decision trees, neural networks, ensemble models, etc.

Module 3

Automating fraud detection process -Fraud detection models using unsupervised analytics hierarchical clustering, non-hierarchical clustering, agglomerative and divisive techniques k-means clustering, self organizing maps, etc.

Module 4

Fraud detection models using social network analytics (homophily, featurization, egonets, PageRank, bigraphs etc. -Verification of results and understand how to prosecute fraud. Fraud detection and prevention-case studies.

References

1. Baesens, B., Van Vlasselaer, V., & Verbeke, W. (2015). *Fraud analytic using descriptive, predictive, and social network techniques: a guide to data science for fraud detection.* John Wiley & Sons.
2. Nigrini, M. J. (2011). F*orensic analytics: methods and techniques for forensic accounting investigations* (Vol. 558). John Wiley & Sons.

PG20DA402E5: Neural Networks and Deep Learning

Objectives: To introduce various concepts like ANN, CNN,SNN, RBN, ELU,RFN,RBM etc for deep learning.

Outcomes: Now students are aware of different types of neural networks and the principles of soft computing. They are now able to carry out deep learning using Python.

Module 1

Neural Networks-Application Scope of Neural Networks- Fundamental Concept of ANN: The Artificial Neural Network-Biological Neural Network-Comparison between Biological Neuron and Artificial Neuron-Evolution of Neural Network. Basic models of ANN-Learning Methods-Activation Functions-Importance Terminologies of ANN.

Module 2

Shallow neural networks- Perceptron Networks-Theory- Perceptron Learning Rule Architecture-Flowchart for training Process- Perceptron Training Algorithm for Single and Multiple Output Classes. Back Propagation Network- Theory- Architecture-Flowchart for training process-Training Algorithm-Learning Factors for Back-Propagation Network. Radial Basis Function Network RBFN: Theory, Architecture, Flowchart and Algorithm.

Module 3

Conventional Neural Networks (CNN) – Introduction – Components of CNN Architecture – Rectified Linear Unit (ReLU) Layer – Exponential Linear Unit (ELU, or SELU) – Unique Properties of CNN –Architectures of CNN –Applications of CNN. Reccurent Neural Network-introduction-The Architecture of Recurrent Neural Network- The Challenges of Training Recurrent Networks- Echo-State Networks- Long Short-Term Memory (LSTM) – Applications of RNN.

Module 4

Auto encoder- Introduction – Features of Auto encoder Types of Auto encoder Restricted Boltzmann Machine- Boltzmann Machine – RBM Architecture –Example – Types of RBM.

References

1. S.N.Sivanandam, S. N. Deepa, *Principles of Soft Computing*, Wiley-India, 3rd Edition, 2018.
2. S Lovelyn Rose, L Ashok Kumar, D Karthika Renuka, *Deep Learning Using Python*, Wiley-India, 1st Edition, 2019.
3. Frank Kane (2018) *Machine Learning, Data Science and Deep Learning with Python*, Manning Publications
4. Charu C. Aggarwal, *Neural Networks and Deep Learning*, Springer, September 2018.
5. Francois Chollet, *Deep Learning with Python*, Manning Publications; 1st edition, 2017
6. John D. Kelleher, *Deep Learning* (MIT Press Essential Knowledge series), The MIT Press, 2018

PG20DA402E6: Bayesian Inference

Objectives: To introduce basic concepts and tools like prior information, posterior information in Bayesian Inference and Computing.

Outcomes: Students have understood basics of Bayesian Inference and are able to apply the computational tools for practical purposes in data science and analytics using prior information and sample data.

Module 1

Introduction: Basics on minimaxity: Bayesian inference, Bayesian estimation, loss function, principle of minimum expected posterior loss, quadratic and other common loss functions, Advantages of being a Bayesian, HPD confidence intervals, testing, credible intervals, prediction of a future observation. Robustness and sensitivity, classes of priors, conjugate class, neighbourhood class, density ratio class, different methods of objective priors: Jeffrey's prior, probability matching prior, conjugate priors and mixtures, posterior robustness: measures and techniques

Module 2

Multiparameter and multivariable models: Basics of decision theory, multi-parameter models, Multivariate models, linear regression, asymptotic approximation to posterior distributions

Module 3

Model selection and hypothesis testing: Selection criteria and testing of hypothesis based on objective probabilities and Bayes' factors, large sample methods: limit of posterior distribution, consistency of posterior distribution, asymptotic normality of posterior distribution.

Module 4

Bayesian computations: Analytic approximation, E- M Algorithm, Monte Carlo sampling, Markov Chain Monte Carlo Methods, Metropolis – Hastings Algorithm, Gibbs sampling, examples, convergence issues

References

1. Bolstad W. M. & Curran, J.M., (2016). Introduction to Bayesian Statistics 3rd Ed. Wiley, New York
2. Christensen, R. J., Branscum, A. W., & Hanson T.E., (2011). Bayesian Ideas and data analysis : A introduction for scientist and Statisticians, Chapman and Hall, London
3. Jim, A., (2009). *Bayesian Computation with R*, second edition, Springer, New York
4. Gelman,A.,Carlin, A.J.B. , Stern, H.S. & Rubin, D.B. , (2004). *Bayesian Data Analysis*, 2nd Ed. Chapman & Hall.

PG20DA402E7: Probabilistic Graphical Models

Objective: To learn about the probabilistic graphical models that has got widespread applicability.

Outcomes: The students are able to understand and use various probabilistic graphical models.

Module 1

Probabilistic reasoning: Representing uncertainty with probabilities – Random variables and joint distributions – Independence – Querying a distribution - Graphs **Representation:** Bayesian Network (BN) representation – Independencies in BN – Factorizing a distribution – D-separation- Algorithm for D-separation – From distributions to Graphs

Module 2

Undirected Graphical Models: Factor products – Gibbs distribution and Markov networks – Markov network independencies – Factor graphs – Learning parameters – Conditional Random Fields

Module 3

Gaussian Network Models: Multivariate Gaussians – Gaussian Bayesian networks – Gaussian Markov Random Fields – Exact Inference: variable elimination- Sum-product and belief updates– The Junction tree algorithm

Module 4

Learning: Learning Graphical Models – Learning as optimization – Learning tasks – Parameter estimation – Structure learning in BN – Learning undirected models – Actions and decisions

References

1. Bellot, D. (2016). *Learning probabilistic graphical models in R*. Packt Publishing Ltd.
2. Luis, E. S. (2015). *Probabilistic Graphical Models, 1e*, Springer Nature.
3. Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.
4. Borgelt, C., Steinbrecher, M., & Kruse, R. R. (2009). *Graphical models: representations for learning, reasoning and data mining*. John Wiley & Sons.

PG20DA402E8: Complex Networks

Objectives: to introduce various concepts to model and visualize network structure and understand its dynamics.

Outcomes: Understand network data and representations; Execute graph algorithms; Carryout basic transformation and visualization; Analyse graph visualization algorithms; Analyse real-world networks.

Module 1

Networks of information – Mathematics of networks – Measures and metrics – Large scale structure of networks – Matrix algorithms and graph partitioning

Module 2

Network models – Random graphs – walks on graphs - Community discovery – Models of network formation – Small world model - Evolution in social networks – Assortative mixingReal networks - Evolution of random network - Watts-Strogatz model – Clustering coefficient - Power Laws and Scale-Free Networks – Hubs - Barabasi-Albert model – measuring preferential attachment- Degree dynamics – non-linear preferential attachment

Module 3

Processes on networks – Percolation and network resilience – Epidemics on networks – Epidemic modelling - Cascading failures – building robustness- Dynamical systems on networks – The Bianconi-Barabási model – fitness measurement – Bose-Einstein condensation

Module 4

Models for social influence analysis – Systems for expert location – Link prediction – privacy analysis – visualization – Data and text mining in social networks - Social tagging Module V Social media - Analytics and predictive models – Information flow – Modelling and prediction of flow -Missing data - Social media datasets – patterns of information attention – linear influence model – Rich interactions

References

1. Barabási, A. L. (2016). *Network science*. Cambridge university press.
2. Aggarwal, C. C., (ed.), (2011). *Social Network Data Analytics*, 1e, Springer.
3. Newman, M. J., (2010). *Networks: An introduction*, 1e, Oxford University Press.
4. Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world.,* 1e, Cambridge University Press.

REGULATIONS OF THE POSTGRADUATE PROGRAMMES UNDER CREDIT SEMESTER SYSTEM-MAC-PG-CSS2020 (2020 Admission onwards)

1. SHORT TITLE

1.1 These Regulations shall be called "Mar Athanasius College (Autonomous) Regulations (2020) governing Postgraduate Programmes under the Credit Semester System (MAC-PG-CSS2020)".

1.2 These Regulations shall come into force from the Academic Year 2020-2021.

2. SCOPE

2.1 The regulations provided herein shall apply to all Regular Postgraduate (PG) Programmes, M.A. /M.Sc. /M.Com. conducted by Mar Athanasius College (Autonomous) with effect from the academic year 2020-2021 admission onwards.

3. DEFINITIONS

3.1 **'Academic Committee'** means the Committee constituted by the Principal under this regulation to monitor the running of the Post-Graduate programmes under the Credit Semester System (MAC-PG-CSS2020).

3.2 **'Academic Week'** is a unit of five working days in which distribution of work is organized from day one to day five, with five contact hours of one hour duration on each day. A sequence of 18 such academic weeks constitutes a semester.

3.3 '**Audit Course**' is a course for which no credits are awarded.

3.4 **'CE' means Continuous Evaluation (Internal Evaluation)**

3.5 **'College Co-ordinator'** means a teacher from the college nominated by the Principal to look into the matters relating to MAC-PG-CSS2020 for programmes conducted in the College.

3.6 **'Comprehensive Viva-Voce'** means the oral examinations conducted by the appointed examiners and shall cover all courses of study undergone by a student for the programme.

3.7 **'Common Course'** is a core course which is included in more than one programme with the same course code.

3.8 '**Core Course**' means a course that the student admitted to a particular programme must successfully complete to receive the Degree and which cannot be substituted by any other course.

3.9 '**Course**' means a segment of subject matter to be covered in a semester. Each Course is to be designed variously under lectures / tutorials / laboratory or fieldwork / seminar / project /practical training / assignments/evaluation etc., to meet effective teaching and learning needs.

3.10 **'Course Code'** means a unique alpha numeric code assigned to each course of a programme.

3.11 '**Course Credit**' One credit of the course is defined as a minimum of one hour lecture /minimum of 2 hours lab/field work per week for 18 weeks in a Semester. The course will be considered as completed only by conducting the final examination.

3.12 **'Course Teacher'** means the teacher of the institution in charge of the course offered in the programme.

3.13 **'Credit (Cr)'** of a course is a numerical value which depicts the measure of the weekly unit of work assigned for that course in a semester.

3.14 '**Credit Point(CP)**' of a course is the value obtained by multiplying the grade point (GP) by the Credit (Cr) of the course **CP=GP x Cr**.

3.15 '**Cumulative Grade Point Average(CGPA)**' is the value obtained by dividing the sum of credit points in all the courses taken by the student for the entire programme by the total number of credits and shall be rounded off to two decimal places. CGPA determines the overall performance of a student at the end of a programme.

(CGPA = Total CP obtained/ Total credits of the programme)

3.16 '**Department**' means any teaching Department offering a programme of study in the institution.

3.17 '**Department Council**' means the body of all teachers of a Department in a College.

3.18 **'Dissertation'** means a long document on a particular subject in connection with the project /research/ field work etc.

3.19 '**Duration of Programme**' means the period of time required for the conduct of the programme. The duration of post-graduate programme shall be 4 semesters spread over two academic years.

3.20 '**Elective Course**' means a course, which can be substituted, by equivalent course from the same subject.

3.21 **'Elective Group'** means a group consisting of elective courses for the programme.

3.22 **'ESE' means End Semester Evaluation (External Evaluation).**

3.23 **'Evaluation'** is the process by which the knowledge acquired by the student is quantified as per the criteria detailed in these regulations.

3.24 **External Examiner** is the teacher appointed from other colleges for the valuation of courses of study undergone by the student in a college. The external examiner shall be appointed by the college.

3.25 '**Faculty Advisor**' is a teacher nominated by a Department Council to coordinate the continuous evaluation and other academic activities undertaken in the Department.

3.26 '**Grace Grade Points**' means grade points awarded to course(s), recognition of the students' meritorious achievements in NSS/ Sports/ Arts and cultural activities etc.

3.27 '**Grade Point**' (GP) Each letter grade is assigned a Grade point (GP) which is an integer indicating the numerical equivalent of the broad level of performance of a student in a course.

3.28 '**Grade Point Average(GPA)**' is an index of the performance of a student in a course. It is obtained by dividing the sum of the weighted grade point obtained in the course by the sum of the weights of Course.**(GPA= ∑WGP / ∑W)**

3.29 '**Improvement Course**' is a course registered by a student for improving his performance in that particular course.

3.30 **'Internal Examiner'** is a teacher nominated by the department concerned to conduct internal evaluation.

3.31 '**Letter Grade**' or '**Grade**' for a course is a letter symbol (A+, A, B+, B, C+, C, D) which indicates the broad level of performance of a student for a course.

3.32 **MAC-PG-CSS2020 means Mar Athanasius College Regulations Governing Post Graduate programmes under Credit Semester System, 2020.**

3.33 '**Parent Department**' means the Department which offers a particular post graduate programme.

3.34 '**Plagiarism**' is the unreferenced use of other authors' material in dissertations and is a serious academic offence.

3.35 '**Programme**' means the entire course of study and Examinations.

3.36 **'Project'** is a core course in a programme. It means a regular project work with stated credits on which the student undergo a project under the supervision of a teacher in the parent department/ any appropriate research centre in order to submit a dissertation on the project work as specified. It allows students to work more autonomously to construct their own learning and culminates in realistic, student-generated products or findings.

3.37 '**Repeat Course**' is a course to complete the programme in an earlier registration.

3.38 '**Semester**' means a term consisting of a minimum of 90 working days, inclusive of examination, distributed over a minimum of 18 weeks of 5 working days each.

3.39 '**Seminar**' means a lecture given by the student on a selected topic and expected to train the student in self-study, collection of relevant matter from various resources, editing, document writing and presentation.

3.40 '**Semester Grade Point Average(SGPA)**' is the value obtained by dividing the sum of credit points (CP) obtained by the student in the various courses taken in a semester by the total number of credits for the course in that semester. The SGPA shall be rounded off to two decimal places. SGPA determines the overall performance of a student at the end of a semester (SGPA = Total CP obtained in the semester / Total Credits for the semester).

3.41 '**Tutorial**' means a class to provide an opportunity to interact with students at their individual level to identify the strength and weakness of individual students.

3.42 '**Weight**' is a numeric measure assigned to the assessment units of various components of a course of study.

3.43 **University** means Mahatma Gandhi University Kottayam to which the college is affiliated.

3.44 '**Weighted Grade Point (WGP)**' is grade points multiplied by weight. **(WGP=GPxW)**

3.45 **'Weighted Grade Point Average (WGPA)' is an index of the performance of a student in a course. It is obtained by dividing the sum of the weighted grade points by the sum of the weights. WGPA shall be obtained for CE (Continuous Evaluation) and ESE (End Semester Evaluation) separately and then the combined WGPA shall be obtained for each course.**

4. ACADEMIC COMMITTEE

4.1. There shall be an Academic Committee constituted by the Principal to Manage and monitor the working of MAC-PG-CSS2020.

4.2. The Committee consists of:

1. Principal
2. Dean, Administration
3. Dean, Academics
4. IQAC Coordinator
5. Controller of Examinations

6. One Faculty each representing Arts, Science, Commerce, Languages, and Self Financing Programmes

5. PROGRAMME STRUCTURE

5.1 Students shall be admitted to post graduate programme under the various Faculties. The programme shall include three types of courses, Core Courses, Elective Courses and Common core courses. There shall be a project with dissertation and comprehensive viva-voce as core courses for all programmes. The programme shall also include assignments / seminars/ practical's etc.

5.2 No regular student shall register for more than 25 credits and less than16 credits per semester unless otherwise specified. The total minimum credits, required for completing a PG programme is 80.

5.3. Elective Courses and Groups

5.3.1 There shall be various groups of Programme Elective courses for a Programme such as Group A, Group B etc. for the choice of students subject to the availability of facility and infrastructure in the institution and the selected group shall be the subject of specialization of the programme.

5.3.2 The elective courses shall be either in fourth semester or distributed among third and fourth semesters. There may be various groups of Elective courses (three elective courses in each group) for a programme such as Group A, Group B etc. for the choice of students, subject to the availability of facility and infrastructure in the institution.

5.3.3 The selection of courses from different elective groups is not permitted.

5.3.4 The elective groups selected for the various Programmes shall be intimated to the Controller of Examinations within two weeks of commencement of the semester in which the elective courses are offered. The elective group selected for the students who are admitted in a particular academic year for various programmes shall not be changed.

5.4 Project Work

5.4.1. Project work shall be completed in accordance with the guidelines given in the curriculum.

5.4.2 Project work shall be carried out under the supervision of a teacher of the department concerned.

5.4.3. A candidate may, however, in certain cases be permitted to work on the project in an Industrial/Research Organization on the recommendation of the supervising teacher.

5.4.4 There shall be an internal assessment and external assessment for the project work.

5.4.5. The Project work shall be evaluated based on the presentation of the project work done by the student, the dissertation submitted and the viva-voce on the project.

5.4.6 The external evaluation of project work shall be conducted by two external examiners from different colleges and an internal examiner from the college concerned.

5.4.7 The final Grade of the project (External) shall be calculated by taking the average of the Weighted Grade Points given by the two external examiners and the internal examiner.

5.5 Assignments: Every student shall submit at least one assignment as an internal component for each course.

5.6 Seminar Lecture: Every PG student shall deliver one seminar lecture as an Internal component for every course with a weightage of two. The seminar lecture is expected to train the student in self-study, collection of relevant matter from the various resources, editing, document writing and presentation.

5.7 Test Papers(Internal): Every PG student shall undergo at least two class tests as an internal component for every course with a weight one each. The best two shall be taken for awarding the grade for class tests.

5.8. No courses shall have more than 5 credits unless otherwise specified.

5.9. **Comprehensive Viva-Voce** -Comprehensive Viva-Voce shall be conducted at the end of fourth semester of the programme and its evaluation shall be conducted by the examiners of the project evaluation.

5.9.1. Comprehensive Viva-Voce shall cover questions from all courses in the Programme.

5.9.2. There shall be an internal assessment and an external assessment for the Comprehensive Viva-Voce.

6. ATTENDANCE

6.1. The minimum requirement of aggregate attendance during a semester for appearing at the end-semester examination shall be 75%. Condonation of shortage of attendance to a maximum of 15 days in a semester subject to a maximum of two times during the whole period of the programme may be granted by the University.

6.2 If a student represents his/her institution, University, State or Nation in Sports, NCC, or Cultural or any other officially sponsored activities such as college union/ university union etc., he/she shall be eligible to claim the attendance for the actual number of days participated subject to a maximum 15 days in a Semester based on the specific recommendations of the Head of the Department or teacher concerned.

6.3 Those who could not register for the examination of a particular semester due to shortage of attendance may repeat the semester along with junior batches, without considering sanctioned strength, subject to the existing University Rules and Clause 7.2.

6.4. A Regular student who has undergone a programme of study under earlier regulation/ Scheme and could not complete the Programme due to shortage of attendance may repeat the semester along with the regular batch subject to the condition that he has to undergo all the examinations of the previous semesters as per the MAC-PG-CSS2020 regulations and conditions specified in 6.3.

6.5 A student who had sufficient attendance and could not register for fourth semester examination can appear for the end semester examination in the subsequent years with the attendance and progress report from the principal.

7. REGISTRATION/ DURATION

7.1 A student shall be permitted to register for the programme at the time of admission.

7.2 A student who registered for the Programme shall complete the Programme within a period of four years from the date of commencement of the programme.

7.3 Students are eligible to pursue studies for additional post graduate degree. They shall be eligible for award of degree only after successful completion of two years (four semesters of study) of college going.

8. ADMISSION

8.1 The admission to all PG programmes shall be done through the Centralised Allotment Process of Mar Athanasius College (Autonomous), Kothamangalam (MAC-PG CAP) as per the rules and regulations prescribed by the affiliating university and the Government of Kerala from time to time.

8.2 The **eligibility criteria for admission** shall be as announced by the Parent University from time to time.

9. ADMISSION REQUIREMENTS

9.1 Candidates for admission to the first semester of the PG programme M.Sc. Data Analytics through CSS shall be required to have passed **B.Sc. Degree with Mathematics/Statistics/ Data Science as a core subject or B. Tech/ B.E. in Computer Science/ IT or Bachelor of Computer Applications, provided the candidate has studied at least two courses in probability/ Statistics at degree level.**

9.2 Students admitted under this programme are governed by the Regulations in force.

10. PROMOTION:

10.1 A student who registers for the end semester examination shall be promoted to the next semester

10.2 A student having 75% attendance and who fails to register for examination of a particular semester will be allowed to register notionally and is promoted to the next semester, provided application for notional registration shall be submitted within 15 days from the commencement of the next semester.

10.3 The medium of Instruction shall be English except programmes under faculty of Language and Literature.

11. EXAMINATIONS

11.1 **End-Semester Examinations**: The examinations shall be at the end of each Semester of three-hour duration for each centralised and practical course.

11.2 Practical examinations shall be conducted at the end of each semester or at the end of even semesters as prescribed in the syllabus of the particular programme. The number of examiners for the practical examinations shall be prescribed by the Board of Studies of the programmes.

11.3 A question paper may contain short answer type/annotation, short essay type questions/problems and long essay type questions. Different types of questions shall have different weightage.

12. EVALUATION AND GRADING

12.1 **Evaluation:** The evaluation scheme for each course shall contain two parts; (a) End Semester Evaluation(ESE) (External Evaluation) and (b) Continuous Evaluation(CE)(Internal Evaluation). 25% weightage shall be given to internal evaluation and the remaining 75% to external evaluation and the ratio and weightage between internal and external is 1:3. Both End Semester Evaluation(ESE) and Continuous Evaluation(CE) shall be carried out using direct grading system.

12.2 **Direct Grading: The direct grading for CE (Internal) and ESE(External Evaluation) shall be based on 6 letter grades (A+, A, B, C, D and E) with numerical values of 5, 4, 3, 2, 1 and 0 respectively.**

12.3 **Grade Point Average (GPA):Internal and External components are separately graded and the combined grade point with weightage 1 for internal and 3 for external shall be applied to calculate the Grade Point**

Average (GPA) of each course. Letter grade shall be assigned to each course based on the categorization provided in 12.16.

12.4 **Internal evaluation:** The internal evaluation shall be based on predetermined transparent system periodic written tests, assignments, seminars, lab skills, records, viva-voce etc.

12.5 Components of lnternal (CE) and External Evaluation (ESE): Grades shall be given to the evaluation of theory / practical / project / comprehensive viva-voce and all internal evaluations are based on the Direct Grading System.

Proper guidelines shall be prepared by the BOS for evaluating the assignment, seminar, practical, project and comprehensive viva-voce within the framework of the regulation.

12.6 There shall be no separate minimum grade point for internal evaluation.

12.7 **The model of the components and its weightages for Continuous Evaluation (CE) and End Semester Evaluation (ESE) are shown in below:**

a) For Theory (CE) (Internal)

	Components	Weightage
i.	Assignment	1
ii.	Seminar	2
iii.	Best Two Test papers	2(1 each)
Total		**5**

(Average grade of the best two papers can be considered. For test paper all the Questions shall be set in such a way that the answers can be awarded A+, A, B, C, D, E grades)

b) For Theory (ESE) (External)

Evaluation is based on the pattern of Question specified **in 12.15.5**

c) For Practical(CE) (Internal)

Components	Weightage
Written / Lab Test	2
Lab Involvement and Record	1
Viva	2
Total	**5**

(The components and weightage of the practical(Internal) can be modified by the concerned BOS without changing the total weightage 5)

d) For Practical(ESE) (External)

Components	Weightage
Written / Lab Test	7
Lab Involvement and Record	3
Viva	5
Total	**15**

(The components and weightage of the practical (External) can be modified by the concerned BOS without changing the total weightage 15)

e) For Project(CE) (Internal)

Components	Weightage
Relevance of the topic and analysis	2
Project content and presentation	2
Project viva	1
Total	**5**

(The components and the weightage of the components of the Project (Internal) can be modified by the concerned BOS without changing the total weightage 5)

f) ForProject(ESE) (External)

Components	Weightage
Relevance of the topic and analysis	3
Project content and presentation	7
Project viva	5
Total	**15**

(The components and the weightage of the components of the Project (External) can be modified by the concerned BOS without changing the total weightage 15)

g) Comprehensive viva-voce (CE) (Internal)

Components	Weightage
Comprehensive viva-voce(all courses from first semester to fourth semester)	5
Total	**5**

(Weightage of the components of the Comprehensive viva-voce(Internal) shall not be modified.)

h)Comprehensive viva-voce (ESE) (External)

Components	Weightage
Comprehensive viva-voce(all courses from first semester to fourth semester)	15
Total	**15**

(Weightage of the components of the Comprehensive viva-voce(External) shall not be modified.)

12.8 **All grade point averages shall be rounded to two digits.**

12.9 To ensure transparency of the evaluation process, the internal assessment grade awarded to the students in each course in a semester shall be published on the notice board at least one week before the commencement of external examination.

12.10 **There shall not be any chance for improvement for Internal Grade.**

12.11 The course teacher and the faculty advisor shall maintain the academic record of each student registered for the course and a copy should be kept in the college for verification for at least two years after the student completes the programme.

12.12 **External Evaluation.** The external examination in theory courses is to be conducted by the College at the end of the semester. The answers may be written in English or Malayalam except those for the Faculty of Languages. The evaluation of the answer scripts shall be done by examiners based on a well-defined scheme of valuation. The external evaluation shall be done immediately after the examination.

12.13 Photocopies of the answer scripts of the external examination shall be made available to the students on request as per the rules prevailing in the University.

12.14 The question paper should be strictly on the basis of model question paper set and directions prescribed by the BOS.

12.15. **Pattern of Questions**

12.15.1 **Questions shall be set to assess knowledge acquired, standard, and application of knowledge, application of knowledge in new situations, critical evaluation of knowledge and the ability to synthesize knowledge. Due weightage shall be given to each module based on content/teaching hours allotted to each module.**

12.15.2 The question setter shall ensure that questions covering all skills are set.

12.15.3 A question paper shall be a judicious mix of short answer type, short essay type /problem solving type and long essay type questions.

12.15.4 The question shall be prepared in such a way that the answers can be awarded A+, A, B, C, D, E grades.

12.15.5 Weight: Different types of questions shall be given different weights to quantify their range as follows:

Sl.No.	Type of Questions	Weight	Number of questions to be answered
1	Short Answer type questions	1	8 out of 10
2	Short essay / problem solving type questions	2	6 out of 8
3	Long Essay Type questions	5	2 out of 4

12.16**Pattern of question for practical**. The pattern of questions for external evaluation of practical shall be prescribed by the Board of Studies.

12.17**DirectGradingSystem**

Direct Grading System based on a 6- point scale is used to evaluate the Internal and External examinations taken by the students for various courses of study.

Grade	Grade point(G)	Grade Range
A+	5	4.50 to 5.00
A	4	4.00 to 4.49
B	3	3.00 to 3.99
C	2	2.00 to 2.99
D	1	0.01 to 1.99
E	0	0.00

12.18**Performance Grading**

Students are graded based on their performance (GPA/SGPA/CGPA) at the Examination on a 7-point scale as detailed below.

Range	Grade	Indicator
4.50 to 5.00	**A+**	**Outstanding**
4.00 to 4.49	**A**	**Excellent**
3.50 to 3.99	**B+**	**Very good**
3.00 to 3.49	**B**	**Good(Average)**
2.50 to 2.99	**C+**	**Fair**
2.00 to 2.49	**C**	**Marginal**
up to 1.99	**D**	**Deficient(Fail)**

12.19 No separate minimum is required for Internal Evaluation for a pass, but a Minimum C grade is required for a pass in an External Evaluation. However, a minimum C grade is required for pass in a Course

12.20 A student who fails to secure a minimum grade for a pass in a course will be permitted to write the examination along with the next batch.

12.21 **Improvement of Course**- The candidate who wish to improve the grade/grade point of the external examination of the of a course/ courses he/ she has passed can do the same by appearing in the external examination of the semester concerned along with the immediate junior batch. This facility is restricted to first and second semester of the programme.

12.22 **One Time Betterment Programme**- A candidate will be permitted to improve the **CGPA** of the programme within a continuous period of four semesters immediately following the completion of the programme allowing only once for a particular semester. The **CGPA** for the betterment appearance will be computed based on the **SGPA** secured in the original or betterment appearance of each semester whichever is higher.

If a candidate opts for the betterment of **CGPA** of a programme, he/she has to appear for the external examination of the entire semester(s) excluding practical /project/comprehensive viva-voce. One time betterment programme is restricted to students who have passed in all courses of the programme at the regular (First appearance)

12.23 **Semester Grade Point Average(SGPA) and Cumulative Grade Point Average** (**CGPA) Calculations.** The SGPA is the ratio of sum of the credit point of all courses taken by a student in a semester to the total credit for that semester. After the successful completion of a semester, Semester Grade Point Average(SGPA) of a student in that semester is calculated using the formula given below.

Semester Grade Point Average -SGPA (S_j) = $\sum(C_i \times G_i) / \sum C_i$

(SGPA= Total credit Points awarded in a semester / Total credits of the semester)

Where 'S_j' is the j^{th} semester, 'G_i' is the grade point scored by the student in the i^{th} course 'C_i' is the credit of the i^{th} course.

12.24 Cumulative Grade Point Average (**CGPA)** of a programme is calculated using the formula:-

Cumulative Grade Point Average (CGPA) = $\sum(C_i \times S_i) / \sum C_i$

(CGPA= Total credit Points awarded in all semester / Total credits of the programme)

Where 'C_i' is the credit for the i^{th} semester, 'S_i' is the SGPA for the i^{th} semester. The **SGPA** and **CGPA** shall be rounded off to 2 decimal points.

For the successful completion of semester, a student shall pass all courses and score a minimum **SGPA** of 2.0. However a student is permitted to move to the next semester irrespective of her/his **SGPA**

13. GRADE CARD

13.1 The Institution under its seal shall issue to the students, a consolidated grade card on completion of the programme, which shall contain the following information.

a) Name of the University.

b) Name of college

c) Title of the PG Programme.

d) Name of Semesters

e) Name and Register Number of students

f) Code, Title, Credits and Max GPA(Internal, External & Total) of each course (theory &practical), project, viva etc in each semester.

g) Internal, external and Total grade, Grade Point (G), Letter grade and Credit point (P) in each course opted in the semester.

h) The total credits and total credit points in each semester.

i) Semester Grade Point Average (SGPA) and corresponding Grade in each semester

j) Cumulative Grade Point Average (CGPA), Grade for the entire programme.

k) Separate Grade card will be issued.

l) Details of description of evaluation process- Grade and Grade Point as well as indicators, calculation methodology of SGPA and CGPA as well as conversion scale shall be shown on the reverse side of the grade card.

14. AWARD OF DEGREE - The successful completion of all the courses with 'C' grade within the stipulated period shall be the minimum requirement for the award of the degree.

15. MONITORING COMMITTEE

There shall be a Monitoring Committee constituted by the Principal to monitor the internal evaluations conducted.

16. RANK CERTIFICATE

Rank certificate shall be issued to candidates who secure positions 1st and 2nd. Candidates shall be ranked in the order of merit based on the CGPA secured by them. Grace grade points awarded to the students shall not be counted for fixing the rank. Rank certificate shall be signed by the Principal and the Controller of Examinations.

17. **GRIEVANCE REDRESSAL COMMITTEE**

17.1 Department level: The College shall form a Grievance Redressal Committee in each Department comprising of the course teacher and one senior teacher as members and the Head of the Department as Chairperson. The Committee shall address all grievances relating to the internal assessment grades of the students.

17.2. College level: There shall be a college level Grievance Redressal Committee comprising of faculty advisor, college co-ordinator, one senior teacher and one staff council member and the Principal as Chairperson.

18. FACTORY VISIT / FIELD WORK/VISIT TO A REPUTED RESEARCH INSTITUTE/ STUDENT INTERACTION WITH RENOWNED ACADEMICIANS may be conducted for all Programmes before the commencement of Semester III.

19. Each student may undertake Internship/on the job training for a period of not less than 15 days. The time, duration and structure of internship/on the job training can be modified by the concerned Board of Studies.

20. **TRANSITORYPROVISION**

Notwithstanding anything contained in these regulations, the Principal shall, for a period of three year from the date of coming into force of these regulations, have the power to provide by order that these regulations shall be applied to any programme with such modifications as may be necessary.

21. **REPEAL**

The Regulations now in force in so far as they are applicable to programmes offered by the college and to the extent they are inconsistent with these regulations are hereby repealed. In the case of any inconsistency between the existing regulations and these regulations relating to the Credit Semester System in their application to any course offered in a College, the latter shall prevail.

22. **Credits allotted for Programmes and Courses**

20.1 Total credit for each programme shall be **80**.

20.2 Semester-wise total credit can vary from 16to25

20.3 The minimum credit of a course is 2 and maximum credit is 5

23. **Common Course:** If a course is included as a common course in more than one programme, its credit shall be same for all programmes.

24. **Course Codes:** The course codes assigned for all courses (Core Courses, Elective Courses, Common Courses etc.) shall be unique.