AC 20.08.22 ITEM NO: 1.1.3

Deccan Education Society's

Kirti M. Doongursee College of Arts, Science and Commerce (AUTONOMOUS)

Affiliated to

UNIVERSITY OF MUMBAI

Syllabus for

Program: Master of Science
Course: M.Sc.
Subject: Computer Science with Specialization in Data Science

Choice Based Credit System (CBCS)
with effect from
Academic Year 2022-2023

PROGRAMME OUTCOME

PROGRAMME SPECIFIC OUTCOMES (PSOs)

On completion of M.Sc. Data Science programme, students will be able:

PO_01: To become a skilled Data Scientist in industry, academia, or government.

PO_02: To use specialised software tools for data storage, analysis and visualization.

PO_03: To independently carry out research/investigation to solve practical problems.

PO_04: To gain problem-solving ability- to assess social issues (ethical, financial, management, analytical and scientific analysis) and engineering problems.

PO_05: To have a clear understanding of professional and ethical responsibility.

PO_06: To collaborate virtually.

PO_07: To have critical thinking and innovative skills.

PO_08: To translate vast data into abstract concepts and to understand database reasoning.

Semester III

Course Code	Course Title	Course Type	Credits
KPSDS22301	Advanced Machine Learning	DSC	4
KPSDS22302	Predictive Modeling and Analytics	DSC	4
KPSDS22303	Data Engineering	DSC	4
Select anyone from the following electives			
KPSDS22304a	Deep Reinforcement Learning	DSE	4
KPSDS22304b	Healthcare Analytics	DSE	
KPSDS22304c	Social Media Analytics	DSE	
KPSDS223P1	Advanced Machine Learning Practical	DSC Practical	2
KPSDS223P2	Predictive Modeling and Analytics Practical	DSC Practical	2
KPSDS223P3	Data Engineering Practical	DSC Practical	2
KPSDS223P4	Research Paper – I	DSE Practical	2
		Total	**24**

M. Sc (Data Science)		Semester –III	
Course Name: Advanced Machine Learning		**Course Code: KPSDS22301**	
Periods per week 1 Period is 60 minutes	**Lectures**	**4**	
	Credits	**4**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	**2½**	**60**
	Theory Internal	**--**	**40**

Objectives	• Understanding Human learning aspects. • Understanding primitives for learnable computers. • Understanding real world problems solved with Advanced Machine Learning.
Learning Outcomes:	CO1: Understand the key issues in Machine Learning and its associated applications in intelligent business and scientific computing. CO2: Acquire the knowledge about different learning models where a learner will be able to explore his skill to generate data base knowledge using the prescribed techniques. CO3: Understand and implement the techniques for extracting the knowledge using advanced machine learning methods. CO4: Achieve adequate perspectives of Advanced Machine learning methods. CO5: Understand the statistical approach related to machine learning. He will also Apply the algorithms to a real-world problem, optimize the models learned and report on the expected accuracy that can be achieved by applying the models.

Pre requisites	Knowledge of Algorithms and mathematical foundation

Unit	Details	Lectures
I	**Introduction:** Machine learning, Examples of Machine Learning Problems, Structure of Learning, learning versus Designing, Training versus Testing, Characteristics of Machine learning tasks, Predictive and descriptive tasks, Machine learning Models: Geometric Models, Logical Models, Probabilistic Models. Features: Feature types, Feature Construction and Transformation, Feature Selection.	**12**
II	A Formal Learning Model, PAC Learning, A More General Learning Model. Learning via Uniform Convergence: Uniform Convergence Is Sufficient for Learnability, Finite Classes Are Agnostic PAC Learnable The VC-Dimension: Infinite-Size Classes Can Be Learnable, The VC-Dimension, Examples, The Fundamental Theorem of PAC learning	**12**
III	**Linear Predictors:** Halfspaces, Linear Regression , Logistic Regression Boosting: Weak Learnability, AdaBoost , Linear Combinations of Base Hypotheses , AdaBoost for Face Recognition Model Selection and Validation: Model Selection Using SRM, Validation, What to Do If Learning Fails Convex Learning Problems: Convexity, Lipschitzness, and Smoothness, Convex Learning Problems, Surrogate Loss Functions	**12**
IV	Rademacher Complexities: The Rademacher Complexity, Rademacher Complexity of Linear Classes, Generalization Bounds for SVM, Generalization Bounds for Predictors with Low Norm Covering Numbers: Covering, From Covering to Rademacher Complexity via Chaining Proof of the Fundamental Theorem of Learning Theory: The Upper Bound for the Agnostic Case, The Lower Bound for the Agnostic Case, The Upper Bound for the Realizable Case Multiclass Learnability: The Natarajan Dimension, The Multiclass Fundamental Theorem, Calculating the Natarajan Dimension, On Good and Bad ERMs	**12**
V	**Probabilistic Model:** Normal Distribution and Its Geometric Interpretations, Naïve Bayes Classifier, Discriminative learning with Maximum likelihood, Probabilistic Models with Hidden variables: Estimation-Maximization Methods, Gaussian Mixtures, and Compression based Models. **Trends In Machine Learning :** Model and Symbols- Bagging and Boosting, Multitask learning, Online learning and Sequence Prediction, Data Streams and Active Learning, Deep Learning, Reinforcement Learning.	**12**

Books and References:					
Sr. No.	**Title**	**Author/s**	**Publisher**	**Edition**	**Year**
01	Machine Learning: The Art and Science of Algorithms that Make Sense of Data	Peter Flach	Cambridge University Press		2012
02	UNDERSTANDING MACHINE LEARNING From Theory to Algorithms	Shai Shalev-Shwartz, Shai Ben-David	Cambridge University Press		2014
03	Introduction to Statistical Machine Learning with Applications in R	Hastie, Tibshirani, Friedman	Springer	2nd	2012
04	Introduction to Machine Learning	Ethem Alpaydin	PHI	2nd	2013

M. Sc (Data Science)		**Semester – III**	
Course Name: Advanced Machine Learning Practical		**Course Code: KPSDS223P1**	
Periods per week (1 Period is 60 minutes)		**4**	
Credits		**2**	
		Hours	**Marks**
Evaluation System	**Practical Examination**	**2**	**50**
	Internal	**--**	**-**

List of Practical:	
	Two Practical Assignments on each unit of the syllabus. Total 10 practical questions to be carried out.

M. Sc (Data Science)	**Semester – III**
Course Name: Predictive Modeling and Analytics	**Course Code: KPSDS22302**

Periods per week 1 Period is 60 minutes	**Lectures**	**4**	
	Credits	**4**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	**2½**	**60**
	Theory Internal	**--**	**40**

Objectives	• Develop an understanding of regression analysis and model building. • Provide the ability to develop relationship between variables • Investigate possible diagnostics in regression techniques • Formulate feasible solution using regression model for real-life problems
Learning Outcomes:	CO1: Develop in-depth understanding of the linear and nonlinear regression model. · demonstrate the knowledge of regression modeling and model selection techniques. CO2: Examine the relationships between dependent and independent variables. CO3: Estimate the parameters and fit a model CO4: Investigate possible diagnostics in regression modeling and analysis. CO5: Validate the model using hypothesis testing and confidence interval approach understand the generalizations of the linear model to binary and count data.
Pre requisites	Knowledge of Algorithms and mathematical foundation

Unit	Details	Lectures
I	**Simple Regression Analysis:** Introduction to a linear and nonlinear model. Ordinary Least Square methods. Simple linear regression model, using simple regression to describe a linear relationship. Fitting a linear trend to time series data, Validating simple regression model using t, F and p test. Developing confidence interval. Precautions in interpreting regression results. **Multiple Regression Analysis:** Concept of Multiple regression model to describe a linear relationship, Assessing the fit of the regression line, inferences from multiple regression analysis, problem of overfitting of a model, comparing two regression model, prediction with multiple regression equation.	**12**

II	**Fitting Curves and Model Adequacy Checking:** Introduction, fitting curvilinear relationship, residual analysis, PRESS statistics, detection and treatment of outliers, lack of fit of the regression model, test of lack of fit, Problem of autocorrelation and heteroscedasticity. Estimation of pure errors from near neighbors. **Transformation techniques:** Introduction, variance stabilizing transformations, transformations to linearize the model, BoxCox methods, transformations on the repressor's variables, Generalized and weighted least squares, Some practical applications.	**12**
III	**Multicollinearity:** Introduction, sources of multicollinearity, effects of multicollinearity. Multicollinearity diagnostics: examination of correlation matrix, variance Inflation factors (VIF), Eigen system analysis of X1X. Methods of dealing with Multicollinearity: collecting additional data, model , re-specification, and ridge regression	**12**
IV	**Generalized Linear Models:** link functions and linear predictors, parameter estimation and inference in the GLM, prediction and estimation with the GLM, Residual Analysis, and concept of over dispersion.	**12**
V	**Model building and Nonlinear Regression:** Variable selection, model building, model misspecification. Model validation techniques: Analysis of model coefficients, and predicted values, data splitting method. Nonlinear regression model, nonlinear least squares, transformation to linear model, parameter estimation in nonlinear system, statistical inference in nonlinear regression. **Contemporary issues:** Research and Analytical problems on various applications of the regression analysis and predictive modeling	**12**

Books and References:

Sr. No.	Title	Author/s	Publisher	Edition	Year
01	Introduction to Linear Regression Analysis	Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining	Wiley India Pvt. Ltd	3rd	2016
02	Applied Regression Analysis	Norman R. Draper, Harry Smith	Wiley India Pvt. Ltd	3rd	2016
03	Applied Multivariate Statistical Analysis	Johnson, R A., Wichern, D. W	PHI learning	2013	2013
04	Applied Regression Modeling	Iain Pardoe	John Wiley and Sons	2nd	2012

M. Sc (Data Science)		Semester – III	
Course Name Predictive Modeling and Analytics		Course Code: KPSDS223P2	
Periods per week (1 Period is 60 minutes)		4	
Credits		2	
		Hours	Marks
Evaluation System	Practical Examination	2	50
	Internal	--	-

List of Practical:	
	Two Practical Assignments on each unit of the syllabus. Total 10 practical questions to be carried out.

M. Sc (Data Science)		Semester – III	
Course Name: Data Engineering		Course Code: KPSDS22303	
Periods per week 1 Period is 60 minutes	Lectures	4	
	Credits	4	
		Hours	Marks
Evaluation System	Theory Examination	2½	60
	Theory Internal	--	40

Objectives	
	• To develop the skills of managing the data with respect to knowledge generation. • Provide the ability to design the data engineering process • To propose the data reliability models • To define how to use Machine learning models

Learning Outcomes:	• CO1: Building the storage system with appropriate data technologies • CO2: designing the data pipelines and data flow • CO3: Processing the data infrastructure • CO4: Investigate possible diagnostics by designing Databases for Reliability, Scalability, and Availability, Understanding Data Operations for Flexibility • CO5: Training and measuring the serving Infrastructure for Machine Learning Models

Pre requisites	Knowledge of database concepts and big data

Unit	Details	Lectures
I	**Selecting Appropriate Storage Technologies:** From Buisness Requirements to Storage Systems, Technical Aspects of Data, Types Of Structure, Schema Design Consideration **Building and Operationalizing Storage Systems:** Cloud SQL, Cloud Spanner, Cloud Bigtable, Cloud Firestore, BigQuery, Cloud Memorystore, Cloud Storage, Unmanaged Databases	**12**
II	**Designing Data Pipelines:** Overview Of Data Pipelines, GCP Pipeline Components, Migrating Hadoop and Spark To GCP **Designing a Data Processing Solution:** Designing Infrastructure, Designing for Distributed Processing, Migrating a Data Warehouse	**12**
III	**Building and Operationalizing Processing Infrastructure:** Provisioning and Adjusting Processing Resources, Monitoring Processing Resources **Designing for Security and Compliance:** Identity and Access Management with Cloud IAM, Using IAM with Storage and Processing Services, Data Security, Ensuring Privacy with the Data Loss Prevention API, Legal Compliance	**12**
IV	**Designing Databases for Reliability, Scalability, and Availability:** Designing Cloud Bigtable Databases for Scalability and Reliability, Designing Cloud Spanner Databases for Scalability and Reliability, Designing BigQuery Databases for Data Warehousing **Understanding Data Operations for Flexibility and Portability:** Catloging and Discovery with Data Catalog, Data Preprocessing with Dataprep, Visualizing with Data Studio, Exploring Data with Cloud Datalab, Orchestrating Workflows with Cloud Composer **Deploying Machine Learning Pipelines:** Structure of ML Pipelines, GCP Options for Deploying Machine Learning Pipeline	**12**

V	**Choosing Training and Serving Infrastructure:** Hardware Accelerators, Distributed and Single Machine Infrastructure, Edge Computing with GCP **Measuring, Monitoring, and Troubleshooting Machine Learning Models:** Three Types of Machine Learning Algorithms, Deep Learning, Engineering Machine Learning Models, Common Sources of Error in Machine Learning Models **Leveraging Prebuilt Models as a Service:** Sight, Conversation, Language, Structured Data	**12**

,

Books and References:

Sr. No.	Title	Author/s	Publisher	Edition	Year
01	Professional Data Engineer	DAN SULLIVAN	Sybex - Wiley	3rd	2020
02	Data Driven Science and Engineering	STEVEN L. BRUNTON, J. NATHAN KUTZ	Cambridge University Press	2nd	2019
03	Data Security in Cloud Computing	Vimal Kumar, Sivadon Chaisiri and Ryan Ko	The Institution of Engineering and Technology		2020
04	Data Engineering on Azure	Vlad Riscutia	Manning Publications		2021

M. Sc (Data Science)		**Semester – III**	
Course Name: Data Engineering Practical		**Course Code: KPSDS223P3**	
Periods per week (1 Period is 60 minutes)		**4**	
Credits		**2**	
		Hours	**Marks**
Evaluation System	**Practical Examination**	**2**	**50**
	Internal	**--**	**-**

List of Practical:	
	Two Practical Assignments on each unit of the syllabus. Total 10 practical questions to be carried out.

M. Sc (Data Science)		**Semester – III**	
Course Name: Deep Reinforcement Learning		**Course Code: KPSDS22304a**	
Periods per week 1 Period is 60 minutes	**Lectures**	**4**	
	Credits	**4**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	**2½**	**60**
	Theory Internal	**--**	**40**

Objectives	● To present the mathematical, statistical and computational challenges of building neural networks ● To study the concepts of deep learning ● To enable the students to know deep learning techniques to support real-time applications
Learning Outcomes:	At the end of successful completion of the course the student will be able to: ● CO1: Describes basics of mathematical foundation that will help the learner to understand the concepts of Deep Learning. ● CO2: Understand and describe model of deep learning ● CO3: Design and implement various deep supervised learning architectures for text & image data. ● CO4: Design and implement various deep learning models and architectures. ● CO5: Fundamentals of Reinforcement learning

Pre requisites	Knowledge of Machine learning Algorithms and mathematical concepts

Unit	Details	Lectures
I	**Applied Math and Machine Learning Basics:** Linear Algebra: Scalars, Vectors, Matrices and Tensors , Multiplying Matrices and Vectors , Identity and Inverse Matrices, Linear Dependence and Span, norms, special matrices and vectors, eigen decompositions. **Machine Learning Basics:** Learning Algorithms, Capacity, Overfitting and Underfitting, Hyperparameters and Validation Sets, Estimators, Bias and Variance, Maximum Likelihood Estimation, Bayesian Statistics, Supervised Learning Algorithms, Unsupervised Learning Algorithms, Stochastic Gradient Descent, building a Machine Learning Algorithm, Challenges Motivating Deep Learning	12
II	**Deep Networks:** Deep feedforward network , regularization for deep learning , Optimization for Training deep models	12
III	**Deep Networks:** Convolutional Networks, Advanced Convolution network, Sequence Modelling, Applications	12
IV	**Deep Learning Research:** Linear Factor Models, Autoencoders	12
V	**Fundamentals of Reinforcement Learning:** introduction, reinforcement learning as MDP, learnable functions in reinforcement learning, deep reinforcement learning algorithms, deep learning for reinforcement , reinforcement learning and supervised learning.	12

,

Books and References:

Sr. No.	Title	Author/s	Publisher	Edition	Year
01	Deep Learning	Ian Goodfellow, Yoshua Bengio, Aaron Courvile	An MIT Press book	1st	2016
02	Fundamentals of Deep Learning	Nikhil Buduma	O'Reilly	1st	2017
03	Deep Learning: Methods and Applications	Deng & Yu	Now Publishers	1st	2013
04	Deep Learning CookBook	Douwe Osinga	O'Reilly	1st	2017

M. Sc (Data Science)		Semester –III	
Course Name: Healthcare Analytics		**Course Code: KPSDS22304b**	
Periods per week 1 Period is 60 minutes	**Lectures**	**4**	
	Credits	**4**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	**2½**	**60**
	Theory Internal	**--**	**40**

Objectives	To empower healthcare providers with effective analytical methods and tools that enable and assist them.
Learning Outcomes:	• To understand biomedical and health informatics • To understand healthcare delivery systems, analyze physiological signals from patient monitoring systems. • To understand predictive modeling and its applications to a broad variety of clinical and translational projects. • To understand predictive usage within radiation oncology and disease modeling for sepsis. • To understand dealing with physicians–patient interactions, Insurance claims, and the role of social media in healthcare.

Unit	Details	Lectures
I	**Recent Development in Methodology for Gene Network Problems and Inferences** Introduction, Background, Genetic Data Available, Methodology, Search Algorithm, PC Algorithm, Application/Case Studies **Biomedical Analytics and Morphoproteomics: An Integrative Approach for Medical Decision Making for Recurrent or Refractory Cancers** Introduction, Backgrounds, Methodology, Case Studies **Characterization and Monitoring of Nonlinear Dynamics and Chaos in Complex Physiological Systems** Introduction, Backgrounds, Sensor-Based Chracterization and Modeling of Nonlinear Dynamics, HealthCare Application **Statistical Modeling of Electrocardiography Signal for Subject Monitoring and Diagnosis** Introduction, Basic Elements of ECG , Statistical Modeling of ECG for Disease Diagnosis, Detection of Obstructive Sleep Apnea from Single ECG Lead, Materials And Methods, Results.	12

II	**Modeling and Simulation of Measurement Uncertainty in Clinical Laboratories** Introduction, Background and Literature Review, Model Development Guidelines, Implementations of Guidelines: Enze Assay Uncertainty Model **Predictive Analytics: Classification in Medicine and Biology** Introduction, Background , Machine Learning with Discrete Support Vector Machine Predictive Models , Applying DAMIP to real World Application **Predictive Modeling in Radiation Oncology** Introduction, Predictive Modeling Techniques, Review of Recent Predictive Modeling Application in Radiation Oncology, Modeling Pathologic Response of Esophageal cancer to Chemoradiotherapy, Modeling Clinical Complications after Radiation Therapy , Modeling Tumor Motion with Respiratory Surrogates	**12**
III	**Mathematical Modeling of Innate Immunity Responses of Sepsis: Modeling and Computational Studies.** Background, System Dynamic Mathematics Model (SDMM), Pathogen Strain Selection, Mathematical Models of Innate Immunity of AIR, Discussion. **Systems Analytics: Modeling and Optimizing Clinic Workflow and Patient Care.** Introduction , Background , Challenges and Objectives, Methods and Design to Study , Computational Results, Implementation and ED Performance Comparison, Benefits and Impacts, Scientific Advances **A Multiobjective Simulation Optimization of the Macrolevel Patient Flow Distribution.** Introduction , Literature Review, Problem Description and Modeling , Methodology, Case Study: Adjusting Patient Flow for a Two-Level Healthcare System Centered on the Puth. **Analysis of Resource Intensive Activity Volumes in US Hospitals** Introduction, Structural Classification of Hospitals, ductivity Analysis of Hospitals, Resource and Activity Database for US Hospitals, Activity Based Modeling of Hospitals Operations , Resource use Profile of Hospitals from HUC Activity Data.	**12**
IV	**Discrete-Event Simulation for Primary Care Redesign: Review and a Case Study.** Introduction, Review of Relevant Literature, A Simulation Case Study at a Pediatric Clinic, What-If Analyses. **Temporal and Spatiotemporal Models for Ambulance Demand.** Introduction, Temporal Ambulance Demand Estimation, Spatiotemporal Ambulance Demand Estimation. **Mathematical Optimization and Simulation Analyses for Optimal Liver Allocation Boundaries.** Introduction, Methods, Results, **Predictive Analytics in 30-Day Hospital Readmissions for Heart Failure Patients.** Introduction, Analytics in Prediction Hospital Readmission Risk, Analytics in Recommending Intervention Strategies, Related Work.	**12**

V	**Heterogeneous Sensing and Predictive Modeling of Postoperative Outcomes.** Introduction, Research Background, Research Methodology, Materials and Experimental Design. **Analyzing Patient–Physician Interaction in Consultation for Shared Decision Making.** Introduction, Literature Review, Recent Data Mining Studies , Future Directions. **The History and Modern Applications of Insurance Claims Data in Healthcare Research.** Introduction, Healthcare Cost Predictions, Measuring Quality of Care. **Understanding the Role of Social Media in Healthcare via Analytics: a Health Plan Perspective.** Introduction, Literature Review, Case Study Description, Research Methods and Analytics Tools.	**12**

,

Books and References:

Sr. No.	Title	Author/s	Publisher	Edition	Year
01	HealthCare Analytics from Data to Knowledge to Healthcare Improvement	Hui Yang, Eva K Lee	Wiley		2016
02	Analytics in Healthcare_ A Practical Introduction	Christo El Morr Hossam Ali-Hassan	Springer		2019
03	Machine Learning and AI for Healthcare_ Big Data for Improved Health Outcomes	Arjun Panesar	Apress		2019

M. Sc (Data Science)		**Semester –III**	
Course Name: Social Media Analytics		**Course Code: KPSDS22304c**	
Periods per week 1 Period is 60 minutes	**Lectures**	**4**	
	Credits	**4**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	**2½**	**60**
	Theory Internal	**--**	**40**

Objectives	• To understand all the different parts of a problem and then be able to find improvement points from facts in the past, and to predict the future outcome of present decisions.
Learning Outcomes:	• To understand and deal with any social media network, strategy, or campaign. • Social media analytics integrates with and affects other areas of business. • To give real-world context and insight. • To present decisions. • To learn and think in the field and reach a point where we can effortlessly approach any project with a sharp analytical mind.

Unit	Details	Lectures
I	**Users: The Who of social media.** Measuring Variations in User Behavior in Wikipedia, Long Tails Everywhere: The 80/20 Rule (p/q Rule), Online Behavior on Twitter. **Networks: The How of Social Media.** Types and Properties of Social Networks, Visualizing Networks, Degrees: The Winner Takes All, Capturing Correlations: Triangles, Clustering, and Assortativity. **Temporal Processes: The When of Social Media**. What Traditional Models Tell You About Events in Time, Inter-Event, Bursty Activities of Individuals, Forecasting Metrics in Time.	**12**
II	**Content: The What of Social Media.** Defining Content: Focus on Text and Unstructured Data, Using Content Features to Identify Topics, Extracting Low-Dimensional Information from High-Dimensional Text. **Processing Large Datasets.** MapReduce: Structuring Parallel and Sequential Operations, Multi-Stage MapReduce Flows, Patterns in MapReduce Programming, Sampling and Approximations: Getting Results with Less Computation, Sampling and Approximations: Getting Results with Less Computation, Bloom Filter, Count-Min Sketch, Executing on a Hadoop Cluster (Amazon EC2).	**12**
III	**Learn, Map, and Recommend.** Social Media Services Online, Problem Formulation, Learning and Mapping, Prediction and Recommendation. Social Media Data, From Data to Insights, Luis Madureira, Analytics in Social Media, Dedicated vs. Hybrid Tools.	**12**
IV	Alexander and Frederik Peiniger, Social Network Landscape, Tam Su, The Analytics Process, Armando Terribili, Metrics, Dashboards.	**12**
V	Reports, Milan Veverka, Strategy, Tactics , Michael Wu, Prescriptive Analytics, The Future of Social Media Analytics.	**12**

,**Books and References:**					
Sr. No.	**Title**	**Author/s**	**Publisher**	**Edition**	**Year**
01	Social Media Data Mining and Analytics	Gabor Szabo, Gungor Polatkan, Oscar Boykin, Antonios Chalkiopolos	John Wiley , & Sons		2019
02	Social Media Analytics Strategy	Alex Goncalves	Apress		2017

M. Sc (Data Science)		**Semester – III**	
Course Name: Elective Practical		**Course Code: KPSDS223P4**	
Periods per week (1 Period is 60 minutes)		**4**	
Credits		**2**	
		Hours	**Marks**
Evaluation System	**Examination**	**--**	**50**
	Internal	**--**	**-**
A quality research paper should be written under the guidance of the faculty. The paper is expected to be published in UGC Care Listed, Scopus, Web of Science, IEEE and the like journals. Plagiarism should be less than 10%.			

Semester IV

Course Code	Course Title	Course Type	Credits
KPSDS22401	Data Protection	SEC	4
KPSDS22402	Marketing Analytics	DSC	4
KPSDS22403	Internship		6
KPSDS22404	Project: Document and Viva Voce	DSC	6
KPSDS224P1	Research Paper – II		2
KPSDS224P2	Marketing Analytics Practical		2
		Total	24

M. Sc (Data Science)		Semester –IV	
Course Name: Data Protection		**Course Code: KPSDS22401**	
Periods per week 1 Period is 60 minutes	**Lectures**	**4**	
	Credits	**4**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	**2½**	**60**
	Theory Internal	**--**	**40**

Objectives	To understand data protection and various cases related to it around the world.
Learning Outcomes:	• To get the idea of data protection and laws related to data protection around the world. • To understand jurisdictional issues. • To understand privacy by design and engineering privacy into the Internet. • To understand the concepts of Anonymity and pseudonymity. • To use privacy and data protection as a business opportunity.

Unit	Details	Lectures
I	Mind the Air Group, Europe versus Facebook: An Imbroglio of EU Data Protection Issues, The Context-Dependence of Citizens' Attitudes and Preferences Regarding Privacy and Security, On Locational Privacy in the Absence of Anonymous Payments.	**12**
II	Development Towards a Learning Health System—Experiences with the Privacy Protection Model of the TRANSFORM Project, Could the CE Marking Be Relevant to Enforce Privacy by Design in the Internet of Things? Visions of Technology, Privacy and Innovation: From Disruption to Opportunities	**12**
III	Behavioural Advertising and the New 'EU Cookie Law's as a Victim of Business Resistance and a Lack of Official Determination, Forget About Being Forgotten, Do-It-Yourself Data Protection—Empowerment or Burden?	**12**
IV	Privacy Failures as Systems Failures: A Privacy-Specific Formal System Model, A Precautionary Approach to Big Data Privacy, The Impact of Domestic Robots on Privacy and Data Protection, and the Troubles with Legal Regulation by Design	**12**

V	Is the Human Rights Framework Still Fit for the Big Data Era? A Discussion of the ECTHR's Case Law on Privacy Violations Arising from Surveillance Activities, Metadata, Traffic Data, Communications Data, Service Use Information… What Is the Difference? Does the Difference Matter? An Interdisciplinary View from the UK, Global Views on Internet Jurisdiction and Trans-border Access.	**12**

Books and References:

Sr. No.	Title	Author/s	Publisher	Edition	Year
01	Data Protection on the Move	Serge Gutwirth , Ronald Leenes, Paul De Hert.	Springer		2016
02	Data Protection Act	UK Govt.	Uk Govt.		2018
03	IT Governance	Alan Calder, Steve Watkins	Kogan Page	6th	2015

M. Sc (Data Science)		**Semester – IV**	
Course Name: Research Paper – II		**Course Code: KPSDS224P1**	
Periods per week (1 Period is 60 minutes)		**4**	
Credits		**2**	
		Hours	**Marks**
Evaluation System	**Practical Examination**	**--**	**50**
	Internal	**--**	**-**
A quality research paper should be written under the guidance of the faculty. The paper is expected to be published in UGC Care Listed, Scopus, Web of Science, IEEE and the like journals. Plagiarism should be less than 10%.			

M. Sc (Data Science)	**Semester –IV**
Course Name: Marketing Analytics	**Course Code: KPSDS22402**

Periods per week 1 Period is 60 minutes	**Lectures**	**4**	
	Credits	**4**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	**2½**	**60**
	Theory Internal	**--**	**40**

Objectives	To understand and apply marketing analytics to different real-world scenarios.
Learning Outcomes:	At the end of the course the students should ● apply their understanding of utility theory to measure customer preferences ● identify what customers' value in a product, and assess what they are willing to pay for it ● segment customers based on the differences in what they value, using different techniques, including state of the art latent class methods ● determine the most effective target markets, and how to market to those markets efficiently ● design a study that incorporates all of the above

Unit	Details	Lectures
I	**Introduction:** Slicing and Dicing Marketing Data with PivotTables, Using Excel Charts to Summarize Marketing Data, Using Excel Functions to Summarize Marketing Data, Estimating Demand Curves and Using Solver to Optimize Price, Price Bundling, Nonlinear Pricing.	**12**
II	Price Skimming and Sales, Revenue Management, Simple Linear Regression and Correlation, Using Multiple Regression to Forecast Sales, Forecasting in the Presence of Special Events, Modeling Trend and Seasonality, Ratio to Moving Average Forecast Method, Winter's Method, Using Neural Networks to Forecast Sales	**12**
III	Conjoint Analysis, Logistic Regression, Discrete Choice Analysis, Calculating Lifetime Customer Value	**12**

IV	Using Customer Value to Value a Business, Customer Value, Monte Carlo Simulation, and Marketing Decision Making, Allocating Marketing Resources between Customer Acquisition and Retention, Cluster Analysis, Collaborative Filtering, Using Classification Trees for Segmentation, Using S Curves to Forecast Sales of a New Product, The Bass Diffusion Model, Using the Copernican Principle to Predict Duration of Future Sales	**12**
V	Market Basket Analysis and Lift, RFM Analysis and Optimizing Direct Mail Campaigns, Using the SCANPRO Model and Its Variants, Allocating Retail Space and Sales Resources, Forecasting Sales from Few Data Points, Measuring the Effectiveness of Advertising, Media Selection Models ,Pay Per Click (PPC) Online Advertising, Principal Component Analysis (PCA) ,Multidimensional Scaling (MDS), Classification Algorithms: Naive Bayes Classifier and Discriminant Analysis, Analysis of Variance: One-way ANOVA, Analysis of Variance: Two-way ANOVA, Networks ,The Mathematics Behind The Tipping Point ,Viral Marketing ,Text Mining.	**12**

Books and References:

Sr. No.	Title	Author/s	Publisher	Edition	Year
01	Marketing Analytics: Data-Driven Technique with Microsoft excel	Wayne L. Winston	WILEY		2012
02	Analytical Finance Volume 1: The Mathematics of Equity Derivatives, Markets, Risk and Valuation	Jan R. M. Röman	Palgrave Macmillan		2017
03	Analytical Finance Volume 2: The Mathematics of Equity Derivatives, Markets, Risk and Valuation	Jan R. M. Röman	Palgrave Macmillan		2017

M. Sc (Data Science)		**Semester – IV**	
Course Name: Marketing Analytics Practical		**Course Code: KPSDS224P2**	
Periods per week (1 Period is 60 minutes)		**4**	
Credits		**2**	
		Hours	**Marks**
Evaluation System	**Practical Examination**	**2**	**50**

	Internal	**--**	**-**

List of Practical:	
	Two Practical Assignments on each unit of the syllabus. Total 10 practical questions to be carried out.

KPSDS22403: Internship

- Internship should be of 2 to 3 months with 8 to 12 weeks duration.
- A student is expected to find an internship by himself or herself. However, the institution should assist their students in getting internships in good organizations.
- The home institution cannot be taken as the place of internship.
- A student is expected to devote at least 300 hours physically at the organization.
- Internship can be on any topic covered in the syllabus mentioned in the syllabus, not restricted to the specialization.
- Internship can be done, in one of the following, but not restricted to, types of organizations:
- Software development firms
- Hardware/ manufacturing firms
- Any small scale industries, service providers like banks, clinics/ NGOs/professional institutions like that of CA, Advocate etc
- Civic Depts like Ward office/post office/police station/ panchayat.
- Research Centres/ University Depts/ College as research Assistant for research projects or similar capacities.

Guidelines for making Internship Report in Semester –IV

A student is expected to make a report based on the internship he or she has done in an organization. It should contain the following:

- Certificate: A certificate in the prescribed Performa (given in appendix 1) from the organization where the internship was done.
- Evaluation form: The form filled by the supervisor or to whom the intern was reporting, in the prescribed Performa (given in appendix 2).
- Title: A suitable title giving the idea about what work the student has performed during the internship.
- Description of the organization: A small description of 1 to 2 pages on the organization where the student has interned

- Description about the activities done by the section where the intern has worked: A description of 2 to 4 pages about the section or cell of the organization where the intern actually worked. This should give an idea about the type of activity a new employee is expected to do in that section of the organization.
- Description of work allotted and actually done by the intern: A detailed description of the work allotted and actual work performed by the intern during the internship period. Intern may give a weekly report of the work by him or her if needed. It shall be of around 7 to 10 pages.
- Self assessment: A self assessment by the intern on what he or she has learnt during the internship period. It shall contain both technical as well as interpersonal skills learned in the process. It shall be of around 2 to 3 pages.

KPSDS22404: Guidelines for Documentation of Project Proposal in Semester –IV

A Student should submit project implementation report with following details:

- Title: Title of the project (Same as the one proposed and evaluated at the semester II examination).
- Implementation details: A description of how the project has been implemented. It shall be of 2 to 4 pages.
- Experimental set up and results: A detailed explanation on how experiments were conducted, what software used and the results obtained. Details like screenshots, tables and graphs can come here. It shall be of 6 to 10 pages.
- Analysis of the results: A description on what the results means and how they have been arrived at. Different performing measures or statistical tools used etc may be part of this. It shall be of 4 to 6 pages.
- Conclusion: A conclusion of the project performed in terms of its outcome (May be half a page).
- Future enhancement: A small description on what enhancement can be done when more time and resources are available (May be half a page).
- Program code: The program code may be given as an appendix. The proposal may be of around 20 pages (excluding program code), which needs to be signed by the teacher in charge and head of the Department.

A complete Project report of around 100 pages should be submitted.

(Proforma for the certificate for internship in official letterhead)

This is to certify that Mr/Ms_____ of _____College/Institution worked as an intern as part of her M.Sc. programe in Data Science of University of Mumbai. The particulars of internship are given below: Internship starting date: _____ Internship ending date:_____ Actual number of days worked:_____ Tentative number of hours worked:_____ Hours Broad area of work:_____ A small description of work done by the intern during the period:

_____

_____

_____

_____

_____

_____

Signature: Name: Designation: Contact number: Email: (seal of the organization)

Appendix 2

(Proforma for the Evaluation of the intern by the supervisor/to whom the intern was reporting in the organization)

Professional Evaluation of intern

Name of intern:_____

College/institution:_____

[Note: Give a score in the 1-5 scale by putting √ in the respective cells]

Sr No	Particular	Excellent	Very Good	Good	Moderate	Satisfactory
1	Attendance					
2	Punctuality					
3	Adaptability					
4	Ability to shoulder responsibility					
5	Ability to work in a team					
6	Written and oral communication skills					
7	Problem solving skills					
8	Ability to grasp new concepts					
9	Ability to complete task					
10	Quality of work done					

Evaluation Scheme for First Year (PG) under AUTONOMY

I. Internal Evaluation for Theory Courses – 40 Marks

Continuous Internal Assessment 1 (Class Test, Assignment-Tutorial, Project Presentations) – 40 Marks

II. External Examination for Theory Courses – 60 Marks

Duration: 2 Hours
Theory question paper pattern:

All questions are compulsory.

Question	Based on	Options	Marks
Q.1	Unit I	*Any 3 out of 6*	12
Q.2	Unit II	*Any 3 out of 6*	12
Q.3	Unit III	*Any 3 out of 6*	12
Q.4	Unit IV	*Any 3 out of 6*	12
Q.5	Unit V	*Any 3 out of 6*	12

- All questions shall be compulsory with internal choice within the questions.
- Each Question may be sub-divided into sub questions as a, b, c, d, etc. & the allocation of Marks depends on the weightage of the topic.

III. Practical Examination

- Each core subject carries 50 Marks (30 marks External + 20 marks Internal)
- Duration: 2 Hours for each practical course.
- Minimum 80% practical from each core subjects are required to be completed.
- Certified Journal is compulsory for appearing at the time of Practical Exam