Dr. Bhimrao Ambedkar University, Agra

A State University of Uttar Pradesh (Paliwal Park, Agra -282004)

www.dbrau.ac.in

A Documentary Support

for

Matric No. – 1.1.2

employability/ entrepreneurship/ skill development

under the

Criteria – I

(Curriculum Design and Development)

Key Indicator - 1.1

in

Matric No. – 1.1.2

MASTER OF SCIENCE(DATA SCIENCE)

2022

Mapping of course to:

Employability | Entrepreneurship | Skills Development

Registrar
Dr. B.R.A. University, Agra

Signature Not Verified
RAJEEV KUMAR
Digitally Signed By Registrar

Program: M.Sc. (Data Science)

(Designed under NEP 2020 and based on CBCS)

(Implemented from session 2022-23)

Duration: 4 Semesters/2 years **Total Credits:** 100

Minimum Credits: I&II Sem.(first year): 52 credits + III&IV Sem. (second year): 48 credits= 100 credits

First year Exit: Min 52 credits: **B.Sc.(Research)**

Second year completion: Min 100 credits: **M.Sc.(Data Science)**

Eligibility:

B. Tech. / B.E. / B. Arch. with 55% marks in aggregate OR B.A. / B.Sc. / BCA / BIT from recognized University with 55% marks in aggregate and must have studied Statistics / Mathematics at the qualifying examination.

Definition (Semester): Six Months: 90 Days= 90(days)/6(days)=15 weeks of teaching

1 credit theory=15 hours of teaching in a semester

1 credit practical=30 hours of teaching in a semester

Semester I

Major Courses

S.No.	*Course No.*	*Course Title*	*Credits*	*L:T:P*	*Internal*	*External*	*Teaching Hours*
1	MDS-101	Descriptive Statistics for Data Science	4	3:1:0	25	75	60 hours
2	MDS-102	Linear Algebra and Matrix Computation	4	3:1:0	25	75	60 hours
3	MDS-103	Regression Analysis and Predictive Modelling	4	3:1:0	25	75	60 hours
4	MDS-104	Probability and Distribution Theory	4	3:1:0	25	75	60 hours
5	MDS-105	Lab Work based on 101,102,103,104	4	0:0:4	---	100	120 hours
Minor Elective: Other Faculty (to be completed in I or II semester)							
6		From the available list	4/5/6	3:1:0	25	75	60 hours/ 75 hours/ 90 hours
Value added course/Offered to other faculties as Minor Elective							
7	STAT-404	Data Analysis using R	()4	2:0:2	25	75	60 hours
Major: Industrial Training/Survey/Research Project							
8	MDS-RP01	Industry Based Research Project-I	4		100	---	120 hours

Semester II

Major Courses

S.No.	Course No.	Course Title	Credits	L:T:P	Internal	External	Teaching Hours
1	MDS-201	Programming for Data Science With R	4	3:1:0	25	75	60 hours
2	MDS-202	Fundamental of Data Base Management System	4	3:1:0	25	75	60 hours
3	MDS-203	Bayesian Data Analysis	4	3:1:0	25	75	60 hours
4	MDS-204	Machine Learning	4	3:1:0	25	75	60 hours
5	MDS-205	Lab Work based on 201,202,203,204	4	0:0:4	---	100	120 hours
Minor Elective: Other Faculty (to be completed in I or II semester)							
6			4/5/7	3:1:0	25	75	60 hours/ 75 hours/ 90 hours
Major: Industrial Training/Survey/Research Project							
8	MDS-RP02	Industry Based Research Project-II	4		---	100	120 hours

***Minor Elective**

1. Social Problems in India: Faculty of Arts
2. Sociology of population Studies: Faculty of Arts
3. NGO Management: Faculty of Arts
4. Disaster Management: Faculty of Arts

Note: The minor elective course chosen by the student will be decided by the department to be offered either in I semester or in the II Semester.

Semester III

Major Courses

S.No.	Course No.	Course Title	Credits	L:T:P	Internal	External	Teaching Hours
1	MDS-301	Multivariate Methods in Data Science	4	3:1:0	25	75	60 hours
2	MDS-302	Programming for Data Science with Python	4	3:1:0	25	75	60 hours
3	MDS-303	Operations Research	4	3:1:0	25	75	60 hours
4	MDS-304	Time Series Analysis	4	3:1:0	25	75	60 hours
5	MDS-305	Lab Work based on 301,302,303,304	4	0:0:4	---	100	120 hours
Major: Industrial Training/Survey/Research Project							
8	MDS-RP03	Industry Based Research Project-III	4		100	---	120 hours

Semester IV

Major Courses

S.No.	Course No.	Course Title	Credits	L:T:P	Internal	External	Teaching Hours
1	MDS-401	Big Data Analysis	4	3:1:0	25	75	60 hours
2	MDS-402	Marketing Research and Analysis	4	3:1:0	25	75	60 hours
3	MDS-403	Spatial Statistics for Remotely Sensed Images	4	3:1:0	25	75	60 hours
4	MDS-404	Cloud Computing	4	3:1:0	25	75	60 hours
5	MDS-405	Lab Work based on 401,402,403,404	4	0:0:4	---	100	120 hours
Major: Industrial Training/Survey/Research Project							
8	MDS-RP04	Industry Based Research Project-IV	4		---	100	120 hours

Summary

S.No.	Semester	Major Courses	Minor Elective(other faculty)	Industrial Training/Survey/Research Project	Total
1	Sem. I	20	5/6/7	4	Min 52
2	Sem. II	20		4	
3	Sem. III	20		4	Min 48
4	Sem. IV	20		4	
	Total				100

Department of Statistics
Institute of Social Sciences

M.Sc. (Data Science)	
Program Educational Objectives (PEO's)	The primary objective of the MSc. program in Data Science is to develop skilled professional workforce that is prepared to address the increasing needs in the rapidly expanding area of big data analytics. The program aims to provide skills in quantitative data analyses, data mining, data modeling and prediction, data storage and management, big data processing, data visualization, multimedia big data, programming and communication skills. Software based courses/ training and a large number of practical case studies have been integrated in the program to boost the learner confidence and market acceptability.
Program Outcomes (PO's) :	On completion of M.Sc. Data Science programme, graduates will be able to • Become a skilled Data Scientist in industry, academia, or government organizations. • Use specialist software tools for data storage, analysis and visualization. • Independently carry out research/investigation to solve practical problems

Descriptive Statistics for Data Science

Course No. MDS-101

M.Sc. (Data Science): I Semester		
Course Code: MDS 101	**Descriptive Statistics for Data Science**	**Credit: 4** **Max Marks: 60+40 =100**
Course objectives:	To have basic idea about the presentation and analysis of the data.	
Course outcomes:	On successful completion of this course, the students will be able to • Present data in different graphical forms • Formulate the data and draw inference using parametric and non-parametric tests.	
Unit 1	Data types: discrete, continuous, univariate, multivariate, binary, character, factor. Exploratory data analysis: Box plot, Stem and leaf, dot plot, central tendency measures, measures of dispersion	
Unit 2	Measures of association: consistency and independence of data with special reference to attributes. Karl Pearsons's correlation coefficient: simple, partial and multiple correlations. Regression analysis, covariance matrix, correlation matrix	
Unit 3	Formulation of statistical hypothesis. Hypothesis testing Student: t-test: single mean test, independent sample t-test, paired sample t-test. Chi-square test: goodness of fit, test of independence, test for correlation and test for dispersion. Confidence limits and confidenceintervals.	
Unit 4	ANOVA: Introduction of linear models, one way and two way ANOVA. Non-parametric tests: Sign test, signed rank test, Kolmogrov-Smirnov test, Test of independence (run test). Two sampleproblem: Wilcoxon Mann-Whitney test, Median test, Kolmogrov-Smirnov test, run test.	

	Books Recommended: 1. Lander, J.P. (2017): *R for Everyone-Advanced Analytics and Graphics.* Pearson Education. 2. Dalgaard, P. (2008): *Introductory Statistics with R.* Springer. 3. Robinson, A.P. and Hamann, J. D. (2010): *Forest Analytics with R-An Introduction.* Springer. 4. Wolfe, Douglas A. and Schneider, G. (2017): *Intuitive Introductory Statistics.* Springer. 5. Mood, A.M, Graybill F.A and Boes D.C. (2008): *Introduction to the Theory of Statistics.* McGraw Hill, New Delhi. 6. Montgomery D.C and Runger G.C. (2013): *Applied Statistics and Probability for Engineers.* Wiley India, New Delhi.

Linear Algebra and Matrix Computation

Course No. MDS-102

M.Sc. (Data Science): I Semester		
Course Code: MDS 102	**Linear Algebra and Matrix Computation**	**Credit: 4** **Max Marks: 60+40 =100**
Course objectives:	To study the theory of linear algebra in the light of data science.	
Course outcomes:	On successful completion of this course, the students will be able to • Conceptualized the fundamentals of vectors and metrics in linear algebra. • Apply these concepts in the field of data science.	
Unit 1	Vectors and Matrices: Operations on matrices and vectors, special matrices, partitioned matrices, linear and quadratic forms, rank of matrices, determinant of a matrix, inverse matrices and applications in linear models and multivariate methods.	
Unit 2	Vector and Matrix Calculus: Differentiation of scalar with respect to a vector, differentiation of a scalar with respect to a matrix, differentiation of a vector with respect to vector, differentiation of a matrix with respect to scalar, eigen analysis of real symmetric matrices, properties of eigen values, statistical application to principal component analysis, use of eigen analysis in constrainedoptimization.	
Unit 3	Matrix decomposition: QR, LU, LDU, Cholesky decomposition, generalized inverse, solution of linear equations, Hadamard products, Kronecker products and vector operator.	
Unit 4	Applications: Applications to multivariate normal distribution, estimation of its parameters and testing of hypothesis for mean vector and variance covariance matrix, principal componentanalysis, linear discriminant analysis, linear models.	
	Books Recommended: 1. Fieller, N. (2015): *Basics of Matrix Algebra for Statistics with R.* CRC Press. 2. Gentle, J. E. (2017): *Matrix Algebra- Theory, Computations and Applications in Statistics.* Second edition. Springer. 3. Boyd, S. and Vandenberghe, L. (2018): *Introduction to Applied Linear Algebra Vectors.* Matrices and Least Squares, Cambridge University Press. 4. Allaire, G. and Kaber, S. M. (2008): *Numerical Linear Algebra, Texts in Applied Mathematics.* Springer. 5. Hogben, L. (2014). *Handbook of Linear Algebra.* CRC Press/Taylor & Francis Group. 6. Friedberg, S., Insel, A., and Spence, L. (2019).*Linear Algebra.* 5/e, Pearson.	

Regression Analysis and Predictive Modeling

Course No. MDS-103

M.Sc. (Data Science): I Semester		
Course Code: MDS 103	**Regression Analysis and Predictive Modeling**	**Credit: 4** **Max Marks: 60+40 =100**
Course objectives:	To introduce the advanced regression analysis and to study theprediction based modeling.	
Course outcomes:	On successful completion of this course, the students will be able to • Formulate the linear models in the field of data science. • Use these models in real life problems for prediction.	
Unit 1	Linear Models: Simple linear regression, estimating the coefficients, accuracy of the coefficient estimates, model accuracy. Multiple linear regression: regression coefficients, qualitative Predictors, comparison of linear regression with K-Nearest Neighbours	
Unit 2	Generalized linear models: logistic regression, estimating the regression coefficients, multiple logistic regression, logistic regression for >2 response classes, Poisson regression, other generalised linear models.	
Unit 3	Model diagnostics: residuals, comparing models, cross validations: validation set approach, leave-one-out cross-validation, k-fold cross-validation. Bootstrap: validation set approach, leave-one- out, k-fold cross-validation, stepwise variable selection.	
Unit 4	Regularization: subset selection, shrinkage methods, reduction methods, Elastic net, Bayesian shrinkage, non-linear least squares, model selection.	
	Books Recommended: 1. Lander, J.P. (2017). *R for Everyone-Advanced Analytics and Graphics*. Pearson Education. 2. James, G., Witten, D., Hastie, T. and R. Tibshirani (2013). *An Introduction to Statistical Learning: with Applications in R.* Springer. 3. Montgomery, D.C., Peck, E.A. and Vining, G.G. (2016). *Introduction to Linear Regression Analysis.* Third Ed., Wiley India Pvt. Ltd.. 4. Draper, N.R. and Smith, H. (2015):*Applied Regression Analysis.*WILEY India Pvt. Ltd. New Delhi; Third Edition, 2015.	

N. Akhil

Probability and Distribution Theory

Course No. MDS-104

M.Sc. (Data Science): I Semester		
Course Code: MDS 104	**Probability and Distribution Theory**	**Credit: 4** **Max Marks: 60+40 =100**
Course objectives:	To introduce the concepts of probability and distribution theoryusing R	
Course outcomes:	On successful completion of this course, the students will be able to • Apply the results of probability in data science using R. • Understand the concepts of Probability and its distribution using R.	
Unit 1	Empirical definition of Probability, random variable: discrete and continuous. Discrete probability distributions: Uniform, Bernoulli, Binomial, Poisson, Geometric, Negative Binomial. Simulations from distributions in R.	
Unit 2	Continuous probability distributions: Uniform, normal, exponential, gamma, weibull, lognormal. Simulations from distributions in R.	
Unit 3	Sequence of random variables, laws of convergence in probability, Chebychev's inequality,SLLN, WLLN, Central limit theorem, Illustration using the replicate function of R	
Unit 4	Sampling distributions: statistic, standard test statistic: t, chi-square, F. Interrelationships betweent, chi-square and F. Density estimation. Illustrations of sampling distributions using simulation tools.	
	Books Recommended: 1. Hogg, R., Tanis, E. and Zimmerman, D. (2019): *Probability and Statistical Inference.* Pearson Education India. 2. Lander, J.P. (2017):*R for Everyone-Advanced Analytics and Graphics*. Pearson Education. 3. Murdoch, D. and Braun, J. (2020): *A First Course in Statistical Programming with R.* Cambridge University Press. 4. Miller, I. and Miller, M. (2017): *John E. Freund's; Mathematical Statistics*. Pearson. 5. Shanmugam, R. and Chattamvelli, R. (2015): *Statistics for scientists and engineers.* JohnWiley	

Lab Work Based on MDS 101, 102, 103, 104

Course No. MDS-105

M.Sc. (Data Science): I Semester		
Course Code: MDS 105	**Lab Work Based on MDS 101, 102, 103, 104**	**Credit: 4** **Max Marks: 60+40 =100**
Course objectives:	To learn real life/industry applications of theory	
Course outcomes:		

Programming for Data Science with R

Course No. MDS-201

M.Sc. (Data Science): II Semester		
Course Code: MDS 201	**Programming for Data Science with R**	**Credit: 4** **Max Marks: 60+40 =100**
Course objectives:	To Introduce the elementary and advanced concepts of R language.	
Course outcomes:	On successful completion of this course, the students will be able to • Describe statistical modeling using R • Apply these modeling tools in Statistical/Machine learning	
Unit 1	Data exploration and data wrangling: visualization with *ggplot2*, data transformation with *dplyr*, Exploratory Data Analysis, R studio projects, data wrangling, creating tibbles, data import with readr, tidy data with tidyr, relational data with dplyr, strings with stringr, factors with forcats, dates and times with lubridate	
Unit 2	Programming: pipes with magrittr, functions, function arguments, return values, vectors, atomic vectors, recursive vectors, augmented vectors. Iteration with purr, for loops versusfunctionals, the map functions, other patterns of for loops.	
Unit 3	Modeling: model basics with modelr, visualizing models, formulas and model families, model building, many models with purrr and broom, gapminder, creating list-columns, making tidy data with broom, making tidy data with broom.	
Unit 4	Communication: R markdown, basics, code chunks, YAML header, graphics for communication with *gglpot2*, output options, documents, notebooks, presentations, dashboards, websites.	
	Books Recommended: 1. Wickham, H. and Grolemund, G. (2016): *R for Data Science Import, Tidy, Transform, Visualize, and Model Data.* O'Reilly. 2. Lander, J.P. (2017): *R for Everyone-Advanced Analytics and Graphics*. Pearson Education. 3. Wickham, H. (2014): *Advanced R.* CRC Press. Gandrud, C. (2020): *Reproducible Research with R and R Studio.* 3rd edition, CRC Press.	

Fundamental of Data Base Management System

Course No. MDS-202

M.Sc. (Data Science): II Semester		
Course Code: MDS 202	Fundamental of Data Base Management System	Credit: 4 Max Marks: 60+40 =100
Course objectives:	To Provide Knowledge of data base management through R.	
Course outcomes:	On successful completion of this course, the students will be able to • Know the objectives of data management • To extract data from various data bases	
Unit 1	RDBAM and SQL RDBMS overview, queries on one table, joins, self joins, inner-join, outer- join, multiple relations between tables, set operations, aggregate operations, efficient queries, structured query language	
Unit 2	Commands in SQL, data types in SQL, data manipulation and data processing with SQL	
Unit 3	Graphical Practice: graphical excellence, graphical integrity, theory of data graphics: data-ink and graphical redesign, data-ink maximization and graphical design, multifunctioning graphical elements, data density and small multiples, aesthetics and technique in data graphical design.	
Unit 4	Import/export data from various data bases in/from R, understanding features/installation of MYSQL, SQLite, excel, MongoDB, postgre SQL, hive, Hbase, installing R interfaces of popular data bases.	
	Books Recommended: 1. Walkowiak, S. (2016): B*ig Data Analytics with R.* Packt open source. 2. Sumathi, S. and Esakkirajan, S. (2007): *Fundamentals of Relational Database Management Systems.*Springer. 3. Edward R. Tufte (2001):*The Visual Display of Quantitative Information.* Prajapati, V. (2013). *Big Data Analytics with R and Hadoop.* Packt open source	

Bayesian Data Analysis

Course No. MDS-203

M.Sc. (Data Science): II Semester		
Course Code: MDS 203	**Bayesian Data Analysis**	**Credit: 4** **Max Marks: 60+40 =100**
Course objectives:	To introduce the elementary and advanced concepts of Bayesian DataAnalysis.	
Course outcomes:	On successful completion of this course, the students will be able to • Analyse the data through the techniques of Bayesian inference • Apply the Bayesian inference to real life scenario.	
Unit 1	Fundamentals of Bayesian Inference: probability and inference. single-parameter models, introduction to multi-parameter models, asymptotic and connections to non-Bayesian approaches,hierarchical models.	
Unit 2	Fundamentals of Bayesian data analysis, model checking, evaluating, comparing, and expanding models, posterior predictive checking, graphical posterior predictive checks, evaluating,comparing, and expanding models, information criteria and cross-validation, model comparison based on predictive performance, model expansion.	
Unit 3	Bayesian computation: Introduction to Bayesian computation, Markov chain Monte Carlo simulations, Gibbs sampler, Metropolis and Metropolis-Hastings algorithms, convergence ofchains, Effective number of simulation draws, Hamiltonian Monte Carlo, Hamiltonian dynamics for a simple hierarchical model with STAN.	
Unit 4	Regression Models: Bayesian approach to regression models with single multiple predictors. Hierarchical linear models with Bayesian approach and generalized linear models.	
	Books Recommended: 1. Gelman, A., John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari and Donald B.Rubin (2014): *Bayesian Data Analysis*. 3rd Edition. Chapman and Hill. 2. Peter, D. H. (2009): *A First Course in Bayesian Statistical Methods*. Springer-Verlag. NewYork. 3. McElreath, R. (2019): *Statistical Rethinking: A Bayesian Course with Examples in R andStan.* (Chapman & Hall/CRC Texts in Statistical Science). 4. Kruschke, J. (2014):*Doing Bayesian Data Analysis: A Tutorial with R, JAGS and Stan.*Academic Press. 5. Rasmussen, C.E. and Williams, C.K.I. (2006):*Gaussian Processes for MachineLearning*. MIT Press (2006).	

Machine Learning

Course No. MDS-204

M.Sc. (Data Science): II Semester		
Course Code: MDS 204	**Machine Learning**	**Credit: 4** **Max Marks: 60+40 =100**
Course objectives:	To introduce the basis concepts of machine learning	
Course outcomes:	On successful completion of this course, the students will be able to • Describe the concepts of machine learning • Apply the machine learning tools in data science.	
Unit 1	Machine learning: Concept and issues, Supervised versus unsupervised learning, Regression versus Classification problem, Algorithms versus Models, Model training: regression and classification models, , model assessment , bias-variance trade-off, hyper parameter tuning, cross validation, ROC curves.	
Unit 2	Tree based methods: Basics of decision trees, a simple tree, tree entropy and information gain, Trees versus linear models, pros and cons of trees, overfitting, pruning a tree, Trees versus linear models, bagging, random forests, boosting, fitting of classification and regression trees.	
Unit 3	Support vector machines (SVMs): Overview, separating hyperplane, maximal margin classifier, support vector classifier (SVC): linear classification and classification with non-linear decision boundaries, SVM versus SVC, SVM with more than 2 classes: One-versus-One and One-versus- All case, kernel functions.	
Unit 4	Neural Networks: Overview, single and multilayer neural networks, neural networks for regression and classification. kNN classifier and k means clustering as machine learning tools.	
	Books Recommended: 1. Lewis, N.D. (2017): *Machine Learning Made Easy with R: An Intuitive Step by Step Blueprintfor Beginners.* CreateSpace Independent Publishing Platform. 2. Burger, S.V. (2018): *Introduction to Machine Learning with R: Rigorous mathematicalmodeling.* O Reilly. Lantz, B. (2019): *Machine Learning with R: Expert Techniques for Predictive Modeling.*Packt Publications, 3rd edition.	

Lab Work Based on MDS 201, 202, 203, 204

Course No. MDS-205

M.Sc. (Data Science): II Semester		
Course Code: MDS 205	Lab Work Based on MDS 201, 202, 203, 204	Credit: 4 Max Marks: 60+40 =100
Course objectives:	To learn real life/industry applications of theory	
Course outcomes:		

Employability Entrepreneurship Skill Development