AC 01.09.23 ITEM NO: 1.3

Deccan Education Society's

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

Affiliated to

UNIVERSITY OF MUMBAI

Syllabus for
Program: Masters of Science
Course: First Year
Subject: Data Science

Choice Based Credit System (CBCS)
with effect from
Academic Year 2023-2024

PROGRAM OUTCOMES

PO	Description
On completion of M.Sc. Data Science programme, students will be able:	
PO1	To become a skilled Data Scientist in industry, academia, or government.
PO2	To use specialized software tools for data storage, analysis and visualization.
PO3	To independently carry out research/investigation to solve practical problems.
PO4	To gain problem-solving ability- to assess social issues (ethical, financial, management, analytical and scientific analysis) and engineering problems.
PO5	To have a clear understanding of professional and ethical responsibility.
PO6	To collaborate virtually.
PO7	To have critical thinking and innovative skills.
PO8	To translate vast data into abstract concepts and to understand database reasoning.

Deccan Education Society's

Kirti M. Doongursee College

(Autonomous) Proposed

Curriculum as per NEP-2020

Year of implementation- 2023-2024

Name of the Department-Data Science

Semester	Course Code	Course Title	Vertical	Credit
I	K23PSDSMJ111	Fundamentals of Data Science	Major	4
	K23PSDSMJP11	Fundamentals of Data Science – Practical	Major	2
	K23PSDSMJ112	Statistical Methods for Data Science	Major	4
	K23PSDSMJP12	Statistical Methods for Data Science– Practical	Major	2
	K23PSDSE121	Programming Paradigms	Elective	4
	K23PSDSRM131	Research Methodology	Research Methodology	4
II	K23PSDSMJ211	Artificial Intelligence and Machine Learning	Major	4
	K23PSDSMJP21	Artificial Intelligence and Machine Learning – Practical	Major	2
	K23PSDSMJ212	Soft Computing	Major	4
	K23PSDSMJP21	Soft Computing – Practical	Major	2
	K23PSDSE221	Algorithms for Data Science	Elective	4
	K23PSDSFP24	Project Implementation	Field Project	4

Course Code	MAJOR SEM – I - Fundamentals of Data Science	Credits	Lectures /Week
K23PSDSMJ111	PAPER I	4	4

Course Outcomes:

After successful completion of this course, students would be able to

- Recall key concepts and terminologies used in Data Science.
- To understand the basic building blocks of programming Languages.
- Build, and prepare data for use with a variety of statistical methods and models
- Analyze Data using various Visualization techniques

Unit	Topics	No of Lectures
I	**Introduction to Data Science:** ▪ What is Data? Kinds of data: e.g. static, spatial, temporal, text, media, ▪ Introduction to high level programming language + Integrated Development ▪ Environment (IDE) o Describing data: Exploratory Data Analysis (EDA) + Data Visualization - Summaries, aggregation, smoothing, distributions ▪ Data sources: e.g. relational databases, web/API, streaming, Data collection: e.g. sampling, design (observational vs experimental) and its impact on visualization, modeling and generalizability of results **Data analysis/modeling:** o Question/problem formation along with EDA o Introduction to estimation and inference (testing and confidence intervals) including simulation and resampling o Scope of inference o Assessment and selection e.g. training and testing sets	15
II	**Data Curation, Management and Organization-I** ▪ Query languages and operations to specify and transform data (e.g. projection, selection, join, aggregate/group, summarize) ▪ Structured/schema based systems as users and acquirers of data o Relational (SQL) databases, APIs and programmatic access, indexing o XML and XPath, APIs for accessing and querying structured data contained therein ▪ Semi-structured systems as users and acquirers of data	15

	o Access through APIs yielding JSON to be parsed and structured ▪ Unstructured systems in the acquisition and structuring of data o Web Scraping o Text/string parsing/processing to give structure	
III	**Data Curation, Management and Organization-II** ▪ Security and ethical considerations in relation to authenticating and authorizing access to data on remote systems ▪ Software development tools (e.g. github, version control) ▪ Large scale data systems o Paradigms for distributed data storage o Practical access to example systems (e.g. MongoDB, HBase, NoSQL systems) o Amazon Web Services (AWS) provides public data sets in Landsat, genomics, multimedia	**15**
IV	**Introduction to Statistical Models** ▪ Simple Linear Regression ▪ Multiple Linear Regression ▪ Logistic Regression ▪ Review of hypothesis testing, confidence intervals, etc. ▪ Estimation e.g. likelihood principle, Bayes, ▪ Linear models o Regression theory i.e. least-squares: Introduction to estimation principles o Multiple regression ▪ Transformations, model selection ▪ Interactions, indicator variables, ANOVA o Generalized linear models e.g. logistic, etc. ▪ Alternatives to classical regression e.g. trees, smoothing/splines ▪ Introduction to model selection o Regularization, bias/variance tradeoff e.g. parsimony, AIC, BIC o Cross validation Ridge regressions and penalized regression e.g. LASSO	**15**

Textbooks:

- Algorithms for Optimization Mykel J. Kochenderfer, Tim A. Wheeler, The MIT Press 2019.
- Think Julia: How to Think Like a Computer Scientist by Allen B. Downey and Ben Lauwens 1st Edition 2019 O'reilly.

Additional References:

- Decision Making Under Uncertainty: Theory and Application by Mykel J. Kochenderfer MIT Lincoln Laboratory Series 2015.
- Introduction to Algorithms, By Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest and Clifford Stein 3Ed. (International Edition) (MIT Press) 2009.

Course Code	Fundamentals of Data Science – Practical	Credits	Lectures/Week
K23PSDSMJP11	**Paper I**	**2 credits**	**4 lectures**
Course Outcome: - ● Learn and implement different data modeling techniques ● Understand the design skills of models for data visualization ● Utilize EDA, inference and regression techniques ● Apply data visualization in big-data analytics and pre-processing techniques. ● Empowering students with tools and techniques used in data science			
Perform minimum ten practical based on the basic concepts of each programming paradigm covering the entire syllabus.			

Course Code	MAJOR SEM – I - Statistical Methods for Data Science	Credits	Lectures /Week
K23PSDSMJ112	**Paper II**	**4**	**4**

Course Outcomes: After successful completion of this course, students would be able to ● Identify different types of data and their appropriate statistical measures. ● Understand the basic knowledge on data collection and various statistical elementary tools. ● Apply the theoretical discrete probability distributions like binomial, Poisson, etc., in the relevant application areas. ● Analyze problems and to make better decisions for the future in their fields.

Unit	Topics	No of Lectures
I	**Introduction to Applied Statistics**: The Nature of Statistics and Inference, What is “Big Data”?, Statistical Modelling, Statistical Significance Testing and Error Rates, Simple Example of Inference Using a Coin, Statistics Is for Messy Situations, Type I versus Type II Errors, Point Estimates and Confidence Intervals, Variable Types, Sample Size, Statistical Power, and Statistical Significance, The Verdict on Significance Testing, Training versus Test Data.	**15**
II	**Computational Statistics**: Vectors and Matrices, The Inverse of a Matrix, Eigenvalues and Eigenvectors **Means, Correlations, Counts: Drawing Inferences**: Computing z and Related Scores, Statistical Tests, Plotting Normal Distributions, Correlation Coefficients, Evaluating Pearson’s r for Statistical Significance, Spearman’s Rho: A Nonparametric Alternative to Pearson, Tests of Mean Differences, t-Tests for One Sample, Two Sample t-Test, Paired-Samples t-Test, Categorical Data, Binomial Test, Categorical Data Having More Than Two Possibilities.	**15**
III	**Power Analysis and Sample Size Estimation**: Power for t-Tests, Power for One-Way ANOVA, Power for Correlations. **Analysis of Variance**: Fixed Effects, Random Effects, Mixed Models, Introducing the Analysis of Variance (ANOVA), Performing the ANOVA, Random Effects ANOVA and Mixed Models, One-Way Random Effects ANOVA, Simple and Multiple Linear Regression, Simple Linear Regression, Multiple Regression Analysis, Hierarchical Regression, How Forward Regression Works. **Logistic Regression and the Generalized Linear Model**: Logistic Regression, Logistic Regression, Predicting	**15**

	Probabilities, Multiple Logistic Regression, Training Error Rate Versus Test Error Rate.	
IV	**Multivariate Analysis of Variance (MANOVA) and Discriminant Analysis**: Multivariate Tests of Significance, Example of MANOVA, Outliers, Homogeneity of Covariance Matrices, Linear Discriminant Function Analysis, Theory of Discriminant Analysis, Predicting Group Membership, Visualizing Separation **Principal Component Analysis**: Principal Component Analysis Versus Factor Analysis, Properties of Principal Components, Component Scores, How Many Components to Keep?, Exploratory Factor Analysis, Common Factor Analysis Model, Factor Analysis Versus Principal Component Analysis on the Same, Initial Eigenvalues in Factor Analysis, Rotation in Exploratory Factor Analysis, Estimation in Factor Analysis **Cluster Analysis**: k-Means Cluster Analysis, Minimizing Criteria, Example of k-Means Clustering, Hierarchical Cluster Analysis, Why Clustering Is Inherently Subjective, Nonparametric Tests, Mann– Whitney U Test, Kruskal–Wallis Test, Nonparametric Test for Paired Comparisons and Repeated	**15**

Textbooks:

- Univariate, Bivariate, and Multivariate Statistics Using R Daniel J. Denis Wiley 1st 2020
- Practical Data Science Andreas François Vermeulen A Press 1st 2018
- Data Science from Scratch first Principle in python Joel Grus Shroff Publishers 1st 2017

Additional References:

- Experimental Design in Data science with Least Resources N C Das Shroff Publishers 1st 2018

Course Code	Statistical Methods for Data Science – Practical	Credits	Lectures/Week
K23PSDSMJP12	**Paper II**	**2 credits**	**4 lectures**

Course Outcome: - • Describe basics of mathematical foundation that will help the learner to understand the concepts of Deep Learning. • Understand and describe model of deep learning • Design and implement various deep supervised learning architectures for text & image data. • Analyze and implement various deep learning models and architectures. • Apply various deep learning techniques to design efficient algorithms for real-world applications.	
Perform minimum ten practical based on the basic concepts of each programming paradigm covering the entire syllabus.	

<table>
<tr><td>Course Code</td><td>ELECTIVES SEM – I – Programming Paradigms</td><td>Credits</td><td>Lectures /Week</td></tr>
<tr><td>K23PSDSE121</td><td>Paper I</td><td>4</td><td>4</td></tr>
<tr><td colspan="4"></td></tr>
<tr><td colspan="4">Course Outcomes:
After successful completion of this course, students would be able to
● To develop the programming ability of the student.
● Basic concepts to be cleared from different programming paradigms..
● To learn Different approach towards the programming problem.
● To handle the errors and find suitable solutions which can be applied for the business.</td></tr>
<tr><td colspan="4"></td></tr>
</table>

Unit	Topics	No of Lectures
I	Foundations-Language design, why to study programming language, compilation and interpretation, programming environments. Programming language syntax – Specifying syntax: regular expressions and Context-Free grammar (Token and Regular expressions, Context Free grammar, Derivations and parse trees), Scanning (Generating Finite automation, Scanner code, Table-driven scanning, Lexical errors, pragmas), Parsing (Recursive Descent, Writing L1 grammar, Table driven top down parsing, Bottom up parsing, Syntax errors)	15
II	**OBJECT ORIENTATION** Basic concepts: objects, classes, methods, overloading methods, messages inheritance: overriding methods, single inheritance, multiple inheritance Interfaces, encapsulation, polymorphism. **FUNCTIONAL PROGRAMMING** Definition of a function: domain and range, total and partial functions, strict functions. Recursion, Referential transparency, Side effects of functions	15
III	**LOGIC PROGRAMMING** Basic constructs, Facts: queries, existential queries, conjunctive queries and rules. Definition and semantics of a logic program, Recursive programming: Computational model of logic programming, Goal reduction, Negation in logic programming	15
IV	**SCRIPTING LANGUAGE** What is scripting language, Problem domain (Shell languages, Text processing and report generation, Mathematics and statistics, General purpose scripting, Extension languages),	15

	Scripting the world wide web (CGI scripts, Embedded server side script, client side script, Java Applets, XSLT)	

Textbooks:

- Programming Language Pragmatics Michael Scott Morgan Kaufmann 4th Edition 2015
- The Craft of Functional Programming Thompson, Simon. Haskell: AddisonWesley Professional 2nd Edition 2011
- "Foundations of Programming Languages Design & Implementation" RoostaSeyed Cenage learning 3rd Edition 2003

Additional References:

- Programming Languages: Concepts and Constructs Sethi Ravi Pearson Education 3rd Edition 2000

<table>
<tr><td>Course Code</td><td>RESEARCH METHODOLOGY SEM – I - Research Methodology</td><td>Credits</td><td>Lectures /Week</td></tr>
<tr><td>K23PSDSRM131</td><td>Paper I</td><td>4</td><td>4</td></tr>
<tr><td colspan="4"></td></tr>
<tr><td colspan="4">Course Outcomes:
After successful completion of this course, students would be able to
● To learn the application of Research in different Business Sectors
● To be able to conduct business research with an understanding of all the latest theories.
● To develop the ability to explore research techniques used for solving any real world or innovative problem.
● Analyze the data to help the decision makers in innovative Business process</td></tr>
<tr><td colspan="4"></td></tr>
<tr><td>Unit</td><td colspan="2">Topics</td><td>No of Lectures</td></tr>
<tr><td>I</td><td colspan="2">Introduction: Role of Business Research, Information Systems and Knowledge Management, Theory Building, Organization ethics and Issues

Beginning Stages of Research Process: Problem definition, Qualitative research tools, Secondary data research</td><td>15</td></tr>
<tr><td>II</td><td colspan="2">Research Methods and Data Collection: Survey research, communicating with respondents, Observation methods, Experimental research</td><td>15</td></tr>
<tr><td>III</td><td colspan="2">Measurement Concepts, Sampling and Field work: Levels of Scale measurement, attitude measurement, questionnaire design, sampling designs and procedures, determination of sample size.</td><td>15</td></tr>
<tr><td>IV</td><td colspan="2">Data Analysis and Presentation: Editing and Coding, Basic Data Analysis, Univariate Statistical Analysis and Bivariate Statistical analysis and differences between two variables. Multivariate Statistical Analysis.</td><td>15</td></tr>
<tr><td colspan="4">REFERENCE BOOKS-

1. Business Research Methods William, G.Zikmund, B.J, Babin, J.C. Carr, Atanu Adhikari, M.Griffin Cengage 8e 2016
2. Business Analytics, Albright Winston, Cengage 5e 2015
3. Research Methods for Business Students Fifth Edition, Mark Saunders 2011</td></tr>
</table>

<table>
<tr><td>Course Code</td><td colspan="2">MAJOR SEM – II - Artificial Intelligence and Machine Learning</td><td>Credits</td><td>Lectures /Week</td></tr>
<tr><td>K23PSDSMJ211</td><td colspan="2">Paper I</td><td>4</td><td>4</td></tr>
<tr><td colspan="5"></td></tr>
<tr><td colspan="5">Course Outcomes:
After successful completion of this course, students would be able to
● To know the concepts, risk and benefits of AI and ML.
● To Explain the History of Artificial Intelligence and the future of AI
● Apply Machine Learning concept to solve problems
● Compare and contrast the Machine Learning algorithms for effective business solutions.</td></tr>
<tr><td colspan="5"></td></tr>
<tr><td>Unit</td><td colspan="3">Topics</td><td>No of Lectures</td></tr>
<tr><td>I</td><td colspan="3">Introduction to AI:
The AI problems, AI technique, philosophy and development of Artificial intelligence. Minimax algorithm, alpha-beta pruning, stochastic games, Constraint satisfaction problems. Knowledge and Reasoning: Logical agents, Propositional logic, First-order logic, Inference in FoL: forward chaining, backward chaining, resolution, Knowledge representation: Frames, Ontologies, Semantic web and RDF.</td><td>15</td></tr>
<tr><td>II</td><td colspan="3">Introduction to PROLOG:
Facts and predicates, data types, goal finding, backtracking, simple object, compound objects, use of cut and fail predicates, recursion, lists, simple input/output, dynamic database. Machine Learning: 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</td><td>15</td></tr>
<tr><td>III</td><td colspan="3">Classification and Regression:
Classification: Binary Classification- Assessing Classification performance, Class probability Estimation Assessing class probability Estimates, Multiclass Classification.

Regression: Assessing performance of Regression- Error measures, Overfitting- Catalysts for Overfitting, Case study of Polynomial Regression. Theory of Generalization: Effective number of hypothesis, Bounding the Growth function, VC Dimensions, Regularization theory.</td><td>15</td></tr>
</table>

IV	**Linear Models**: Least Squares method, Multivariate Linear Regression, Regularized Regression, Using Least Square regression for Classification. Perceptron, Support Vector Machines, Soft Margin SVM, Obtaining probabilities from Linear classifiers, Kernel methods for non-Linearity. Logic Based and Algebraic Model: Distance Based Models:Neighbours and Examples, Nearest Neighbours Classification, Distance based clustering-K means Algorithm, Hierarchical clustering **Rule Based Models**: Rule learning for subgroup discovery, Association rule mining. Tree Based Models: Decision Trees, Ranking and Probability estimation Trees, Regression trees, Clustering Trees. 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.	15

Textbooks:

- Artificial Intelligence Elaine Rich, Kevin Knight Tata McGraw Hill 3rd 2017
- Machine Learning: The Art and Science of Algorithms that Make Sense of Data Peter Flach Cambridge University Press 1st Edition, 2012
- Introduction to Statistical Machine Learning with Applications in R Hastie, Tibshirani, Friedman Springer 2nd Edition 2012

Additional References:

- Introduction to Machine Learning EthemAlpaydin PHI 2nd Edition 2013

Course Code	Artificial Intelligence and Machine Learning – Practical	Credits	Lectures/Week
K23PSDSMJP21	Paper I	2 credits	4 lectures

Course Outcomes:

- Learn the key issues and concepts in Artificial Intelligence.
- Acquire the knowledge about classification and regression techniques where a learner will be able to explore his skill to generate database knowledge using the prescribed techniques.
- Understand and implement the techniques for extracting the knowledge using machine learning methods and statistical approaches related to machine learning.
- 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 mode
- Achieve adequate perspectives of big data analytics in various applications like recommender systems, social media applications etc.

Perform minimum ten practical based on the basic concepts of each programming paradigm covering the entire syllabus.

Course Code	MAJOR SEM – II - Soft Computing	Credits	Lectures /Week
K23PSDSMJ212	Paper II	4	4

Course Outcomes:

After successful completion of this course, students would be able to

- Memorize the advantages and limitations of soft computing methods.
- Describe the learning process and architectures of neural networks.
- Apply fuzzy logic to model and handle uncertainty in real-world problems.
- Assess the suitability of different soft computing techniques for specific problem domains.

Unit	Topics	No of Lectures
I	**Artificial Neural Network:** Fundamental concepts, Evolution of neural network, basic model of Artificial Neural Network, Important terminologies, McCulloch Pits neuron, linear separability, Hebb network **Supervised Learning Network**: Perceptron networks, Adaline, MAdaline, Backpropogation network, Radial Basis Function, Time Delay Network, Functional Link Networks, Tree Neural Network.	15
II	**UnSupervised Learning Networks**: Fixed weight competitive nets, Kohonen self-organizing feature maps, learning vectors quantization, counter propogation networks, adaptive resonance theory networks. **Associative Memory Networks**: Training algorithm for pattern Association, Autoassociative memory network, hetroassociative memory network, bi-directional associative memory, Hopfield networks, iterative autoassociative memory networks, temporal associative memory networks.	15
III	**Special Networks**: Simulated annealing, Boltzman machine, Gaussian Machine, Cauchy Machine, Probabilistic neural net, cascade correlation network, cognition network, neo-cognition network, cellular neural network, optical neural network Third Generation Neural Networks: Spiking Neural networks, convolutional neural networks, deep learning neural networks, extreme learning machine model. **Introduction to Fuzzy Logic, Classical sets, Fuzzy sets, Classical Relations and Fuzzy Relations**: Cartesian Product of relation, classical relation, fuzzy relations, tolerance and	15

	equivalence relations, non-iterative fuzzy sets. Membership Function: features of the membership functions, fuzzification and methods of membership value assignments. Defuzzification: Lambda-cuts for fuzzy sets, Lambda-cuts for fuzzy relations, Defuzzification methods. Fuzzy Arithmetic and Fuzzy measures: fuzzy arithmetic, fuzzy measures, measures of fuzziness, fuzzy integrals.	
IV	**Genetic Algorithm**: Biological Background, Traditional optimization and search techniques, genetic algorithm and search space, genetic algorithm vs. traditional algorithms, basic terminologies, simple genetic algorithm, general genetic algorithm, operators in genetic algorithm, stopping condition for genetic algorithm flow, constraints in genetic algorithm, problem solving using genetic algorithm, the schema theorem, classification of genetic algorithm, Holland classifier systems, genetic programming, advantages and limitations and applications of genetic algorithm.	**15**

Textbooks:

- Artificial Intelligence and Soft Computing Anandita Das Battacharya SPD 3rd Edition, 2018
- Principles of Soft computing S.N.Sivanandam S.N.Deepa Wiley 3rd Edition, 2019
- Neuro-Fuzzy and Soft Computing J.S.R.Jang, C.T.Sun and E.Mizutani Prentice Hall of India 1st Edition, 2004

Additional References:

- **Neural Networks, Fuzzy Logic and Genetic Algorithms: Synthesis & Applications S.Rajasekaran, G. A. Vijayalakshami Prentice Hall of India 1st** Edition, **2004**
- **Fuzzy Logic with Engineering Applications Timothy J.Ross McGrawHill 1st** Edition, **1997**
- **Genetic Algorithms: Search, Optimization and Machine Learning Davis E.Goldberg Addison Wesley 1st** Edition, **1989**
- Introduction to AI and Expert System Dan W. Patterson Prentice Hall of India 2nd Edition, 2009

Course Code	**Soft Computing – Practical**	**Credits**	**Lectures/Week**
K23PSDSMJP22	**Paper II**	**2 credits**	**4**
Course Outcome:			

• Identify and describe soft computing techniques and their roles in building intelligent machines • Recognize the feasibility of applying a soft computing methodology for a particular problem • Apply fuzzy logic and reasoning to handle uncertainty and solve engineering problems and also Apply neural networks for classification and regression problems • Apply genetic algorithms to combinatorial optimization problems • Evaluate and compare solutions by various soft computing approaches for a given problem.
Perform minimum ten practical based on the basic concepts of each programming paradigm covering the entire syllabus.

Course Code	ELECTIVE SEM – II - Algorithms for Data Science	Credits	Lectures /Week
K23PSDSE221	Paper I	4	4

Course Outcomes:

After successful completion of this course, students would be able to

- Recognize the key algorithmic concepts used in data science.
- Explain the principles and techniques of algorithm design and analysis.
- Apply various regression methods to solve data manipulation and optimization problems.
- Analyze the efficiency and scalability of algorithms for different problem sizes.

Unit	Topics	No of Lectures
I	**Introduction**: What Is Data Science?, Diabetes in America, Authors of the Federalist Papers, Forecasting NASDAQ Stock Prices, Algorithms, Python, R, Terminology and Notation Data Mapping and Data Dictionaries: Data Reduction, Political Contributions, Dictionaries, Tutorial: Big Contributors, Data Reduction, Election Cycle Contributions, Similarity Measures, Computing Similarity Scalable Algorithms and Associative Statistics: Introduction, Associative Statistics, Univariate Observations, Functions, Histogram Construction, Multivariate Data, Computing the Correlation Matrix, Linear Regression, Computing β	15
II	**Hadoop and MapReduce**: Introduction, The Hadoop Ecosystem, Medicare Payments, The Command Line Environment, Programming a MapReduce Algorithm, Using Amazon Web Services Data Visualization: Introduction, Principles of Data Visualization, Making Good Choices, Harnessing the Machine **Linear Regression Methods**: Introduction, The Linear Regression Model, Introduction to R, Large Data Sets and R, Factors, Analysis of Residuals Healthcare Analytics: Introduction, The Behavioral Risk Factor Surveillance System, Diabetes Prevalence and Incidence, Predicting AtRisk Individuals, Identifying At-Risk Individuals, Unusual Demographic Attribute Vectors, Building Neighborhood Sets	15
III	**Cluster Analysis**: Introduction, Hierarchical Agglomerative Clustering, Comparison of States, Hierarchical Clustering of States, The k-Means Algorithm k-Nearest Neighbor Prediction Functions: Introduction, Notation and Terminology, Distance Metrics, The k-Nearest Neighbor Prediction Function, Exponentially Weighted k-Nearest Neighbors, Digit Recognition, Accuracy Assessment, k-Nearest Neighbor Regression, Forecasting the S&P 500, Forecasting by Pattern Recognition, CrossValidation **The Multinomial Naïve Bayes Prediction Function**: Introduction, The Federalist Papers, The Multinomial Naïve	15

	Bayes Prediction Function, Reducing the Federalist Papers, Predicting Authorship of the Disputed Federalist Papers, Customer Segmentation	
IV	**Forecasting**: Introduction, Working with Time, Analytical Methods, Computing ρτ, Drift and Forecasting, Holt-Winters Exponential Forecasting, Regression-Based Forecasting of Stock Prices, Time-Varying Regression Estimators Real-time Analytics: Introduction, Forecasting with a NASDAQ Quotation Stream, Forecasting the Apple Inc. Stream, The Twitter Streaming API, Sentiment Analysis, Sentiment Analysis of Hashtag Groups	**15**

Textbooks:

- Algorithms for Data Science Brian Steele, John Chandler,Swarna Reddy Springer 1st Edition, 2016
- Data Science Algorithms in a Week David Natingga Packt Publishing 1st Edition, 2017

Additional References:

- **Data Science: Theories, models, Algorithms and Analytics SanjivRanjan Das S.R. Das 1st Edition, 2017**

Course Code	**FP SEM – II – PROJECT IMPLEMENTATION**	**Credits**	**Lectures /Week**
K23PSDSFP24		**4**	**4**

Course Outcome:

- To learn the process of project implementation

- To understand the system, submit the proposal and implement the same in the semester–II.
- To propose project implementation as part of the semester–II.
- Experimental setup, analysis of results, comparison with results of related works, conclusion, and prospects will be part of the project implementation.
- To make a project implementation report and appear for a project viva

PROJECT IMPLEMENTATION

Students need to spend around 133 hours for the project implementation, which fetches 4 credits.

Guidelines for Project Implementation in Semester - II

● A student is expected to devote at least 3 to 4 months of effort to the implementation.
● Students should submit a detailed project implementation report at the time of viva.

Guidelines for Documentation of Project Proposal in Semester –II

A student should submit a project implementation report with the following details:
● Title: Title of the project.
● Objective: A detailed objective of the proposal is needed.
● Related works: A detailed survey of the relevant works done by others in the domain. The student is expected to refer to at least 15 recent (last five years) research papers in addition to textbooks and web links in the relevant topic.
● Methodology: A proper and detailed procedure of how to solve the problem discussed. It shall contain the techniques, tools, software, and data to be used.
● Implementation details: A description of how the project has been implemented.
● Experimental setup and results: A detailed explanation of how experiments were conducted, what software was used, and the results obtained. Details like screenshots, tables, and graphs can come here.
● Analysis of the results: A description of what the results mean and how they have been arrived at. Different performing measures or statistical tools used etc may be part of this.
● Conclusion: A conclusion of the project performed in terms of its outcome
● Future enhancement: A small description of what enhancement can be done when more time and resources are available
● Program code: The program code may be given as an appendix.
The project documentation needs to be signed by the teacher in charge and head of the Department. Students should also attach the certified copy of the internal evaluation report (Appendix III) at the time of Project evaluation and viva as part of the University examination.

Evaluation Scheme for First Year (PG) under NEP (4 credits)

I. Internal Evaluation for Theory Courses – 40 Marks

1) Continuous Internal Assessment(CIA) Assignment - Tutorial/ Case Study/ Project / Presentations/ Group Discussion / Ind. Visit. – 20 marks

2) Continuous Internal Assessment(CIA) ONLINE Unit Test – 20 marks

II. External Examination for Theory Courses – 60 Marks

Duration: 2 Hours

Theory question paper pattern:

Question	Based on	Marks
Q.1	Unit I	15
Q.2	Unit II	15
Q.3	Unit III	15
Q.4	Unit IV	15

- All questions shall be compulsory with internal choice within the questions.
- Each Question may be subdivided 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
- Duration: 3 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

NOTE: To pass the examination, attendance is compulsory in both Internal & External (Theory + Practical) Examinations.