JAI HIND COLLEGE

Basantsing Institute of Science & J. T. Lalvani College of Commerce
and Sheila Gopal Raheja College of Management.

Empowered Autonomous

"A" Road, Churchgate, Mumbai - 400 020, India

Affiliated to
University of Mumbai

Master of Science

Program: MSc. in Big Data Analytics

Choice Based Credit System (CBCS) under NEP-2020
with effect from the academic year 2023-2024

Syllabus as approved by Statutory Committees

LOCF Document

CONTENTS

Preamble

The M. Sc. Big Data Analytics programme was started in 2020 with an aim to make the students employable and impart industry oriented training.

The Master of Science in Big Data Analytics program is of a two years full time post graduate program. It comprises of four semesters with total 88 credits. In collaboration with TCS extends to students rigorously knowledge in big data analytics. The course work contains a compulsory core of professional subjects like statistics, machine learning and enabling technologies for data science relevant to all science disciplines. In order to include practical knowledge on the course, subjects and applications like Data Science, analytics, machine learning, business analytics, and programming languages are included in the curriculum of the course.

With a combination of core subjects, Students do internships to acquire industry- relevant training in semester IV which is TCS supported. The program delivers meticulous training in computational techniques and provides proficiency of data analysis. Preparing students with the tools and techniques required to work with and analyze today's increasingly complex data sets in all areas of the sciences.

The objective of this syllabus is to create a pool of technologically savvy, and upon completion of the M.Sc. in Big Data Analytics program, students will be able to acquire and enhance skills in applied statistics, real analysis and numerical analysis, apply for data analytics job opportunities in the domain of predictive analytics, descriptive statistics, etc.

Acquiring Data Mining, Data Infrastructure, Data Visualization, and Decisions Analysis skills which will allow a student to expand his/her service to different areas and enhance his/her Data Analytics knowledge. Big Data Analytics attain cost-effective solutions and improve decision-making power in multiple development areas, including healthcare, manufacturing, education, media, retail, and even real estate. Students will have an opportunity to select from a variety of industries that match their skills and interests. The knowledge gained through this course will help students acquire plethora of opportunities to serve the industries in various capacities like Big Data Analyst, Big Data Manager, IT Systems Analyst, Operations Analyst, Data Engineer, Quantitative Analyst, Project Manager, Data Scientist, etc..

Credit Framework

Types of Courses

Sr. No	Type of Course	Learner Category
1	Major Mandatory	MSc BDA
2	Major Elective	MSc BDA
3	RM	MSc BDA
4	FP	MSc BDA
5	RP	MSc BDA
6	OJT	MSc BDA

Number of Courses and Credits for Part 1

Type of Course	Number offered of each	Credits of each (Theory+Practical)
Major Mandatory	06	04 (3+1)
MOOC Mandatory	01	02 (Theory)
Value Thinking Mandatory	01	02 (Theory)
Major Elective	02	04 (3+1)
RM	01	04 (3+1)
FP	01	04 (3+1)

Number of Courses and Credits for Part 2

Type of Course	Number offered of each	Credits of each (Theory+Practical)
Major Mandatory	03	04 (3+1)
Major Elective	01	04 (3+1)
RP	01	06 (3+1+2)
OJT	01	22

Semester-wise Courses

Semester	Course Code	Course Name	Types of Courses	Credits
I	JPS-BDA-501-MM	Probability & Stochastic Process	Major Mandatory	3+1
I	JPS-BDA-502-MM	Linear Algebra & Linear Programming	Major Mandatory	3+1
I	JPS-BDA-503-MM	Computing for Data Sciences	Major Mandatory	3+1
I	JPS-BDA-504-MM	MOOC	Major Mandatory	2
I	JPS-BDA-505-ME	Database Management	Major Elective	3+1
I	JPS-BDA-506-RM	Statistical Methods	RM	3+1
II	JPS-BDA-551-MM	Enabling Technologies for Data Science-I	Major Mandatory	3+1
II	JPS-BDA-552-MM	Foundations of Data Science	Major Mandatory	3+1
II	JPS-BDA-553-MM	Advanced Statistical Methods	Major Mandatory	3+1
II	JPS-BDA-554-MM	Value Thinking	Major Mandatory	2
II	JPS-BDA-555-ME	Cloud Computing	Major Elective	3+1
II	JPS-BDA-556-FP	Machine Learning-I	FP	3+1
III	JPS-BDA-601-MM	Enabling Technologies for Data Science-II	Major Mandatory	3+1
III	JPS-BDA-602-MM	Machine Learning-II	Major Mandatory	3+1
III	JPS-BDA-603-MM	Exploratory Data Analysis	Major Mandatory	3+1
III	JPS-BDA-604-ME	Introduction to Econometrics and Finance	Major Elective	3+1
III	JPS-BDA-605-RP	Introduction to Bioinformatics	RP	3+1+2
IV	JPS-BDA-651-OJT	Internship Based Project	OJT	22

Learning Outcome-based Approach

The National Higher Education Qualifications Framework (NHEQF) envisages that students must possess the quality and characteristics of the post graduate of a programme of study, including learning outcomes relating to the disciplinary area in the chosen field of learning and generic learning outcomes that are expected to be acquired by a post graduate on completion of the programme of study.

Learning outcomes are important for recognition. It is important to emphasize on it since the important question asked to the student is-What you can do now that you are a post graduate in Big Data Analytics? The prior specification of the intended educational outcomes helps in maximizing educational effectiveness. All the educational activities from designing curriculum, to teaching pedagogies, to methods of evaluation are directed maximally towards the attainment of specific desired goals. The student's learning outcomes refer to the attainment of the particular competencies acquired by the student on completion of the postgraduate Big Data Analytics program.

Post Graduate Attributes

- **Deep discipline knowledge and intellectual breadth**
  Graduates have comprehensive knowledge and understanding of their subject area, the ability to engage with different traditions of thought, and the ability to apply their knowledge in practice including in multi-disciplinary or multi-professional contexts.
- **Creative, critical thinking, scientific reasoning and problem solving**
  Graduates are effective problems-solvers, able to apply critical, creative and evidence-based thinking to conceive innovative responses to future challenges.
- **Teamwork and communication skills**
  Graduates convey ideas and information effectively to a range of audiences for a variety of purposes and contribute in a positive and collaborative manner to achieving common goals.
- **Professionalism and leadership readiness**
  Graduates engage in professional behavior and have the potential to be entrepreneurial and take leadership roles in their chosen occupations or careers and communities.
- **Independent and lifelong learning**
  A capacity to be a self-directed learner and thinker and to study and work independently. Resulting in continuous learning, resilience, confidence, learning transferable and time management skills and an ability to learn independently
- **Research skills**
  They are able to identify a problem, collect informational resources that can help

address the problem, evaluate these resources for quality and relevance and come up with an effective solution to the problem

Program Specific Objectives

1. Develop relevant programming abilities.
2. Demonstrate proficiency with statistical analysis of data with professional statistical software
3. Develop the ability to build and assess data-based model skills in data management.
4. Apply data science concepts and methods to solve problems in real-world contexts and will communicate these solutions effectively
5. Inculcate critical thinking to carry out scientific investigation objectively without being biased with preconceived notions.
6. Equip the student with skills to analyze problems, formulate a hypothesis, evaluate and validate results, and draw reasonable conclusions thereof.
7. Prepare students for pursuing research or careers in industry in data sciences and allied fields
8. Continue to acquire relevant knowledge and skills appropriate to professional activities and demonstrate the highest standards of ethical issues in data sciences.
9. Basic understanding of statistical methods, probability, mathematical foundations, and computing methods relevant to data analytics.
10. Knowledge about storage, organization, and manipulation of structured data along with challenges associated with big data computing.
11. Understanding about the analytics chain beginning with problem identification and translation, followed by model building and validation with the aim of knowledge discovery in the given domain.
12. Applying dimensionality reduction techniques in finding patterns/features/factors in big data.
13. Estimation of various statistics from stored and/or streaming data in the iterative process of model selection and model building.
14. Modeling optimization techniques such as linear programming, non-linear programming, and transportation techniques in various problem domains such as marketing and supply chain management.
15. Interpret analytical models to make better business decisions.

Teaching Learning Process

1. Lectures
Lectures should be designed to provide the learners with interesting and fresh perspectives on the subject matter. Lectures should be interactive in a way that students work with their teachers to get new insights in the subject area, on which they can build their own bridges to higher learning.

2. Discussions
Discussions are critical components of learning, and can be used as a platform for students to be creative and critical with old and new ideas.

3. Experiential learning
Experiential learning is an engaged learning process whereby students "learn by doing" and by reflecting on the experience and enhancing their technical skills through mini projects. In Semester 4, students need to complete the internship based projects.

4. Industry Expert sessions
Expert sessions, where students are learn about emerging trends in the field of data science and interact with industry experts to gain knowledge about the real time implementation of projects.

5. Case Studies:
Case studies, wherever possible, should be encouraged in order to challenge students to find creative solutions of complex problems and various aspects of the knowledge domain concerned.

6. Team Work
Positive collaboration in the form of teamwork is essential for achieving the intended goals within given deadlines. In the process of teamwork, learners will acquire the skills of managing knowledge, acquisition, critical thinking and other collaborative learners, thereby understanding how to incorporate and balance personalities.

7. Excursion trip:
Excursion trips provide opportunities to the learners to test their in-class learning in real-life situations as well as to understand the functional diversity in the learning spaces.

8. Academics-Industries Interface:
The course encourages students for closer interaction with Industries/corporate/research institutes, etc. for the internship and training program.

9. Blended learning
Combine online educational materials and opportunities for interaction online with traditional place-based classroom methods. Use of digital learning tools with more traditional classroom face to face teaching.

10. Flipped Learning/Flipped Classroom Learning
Improve the student performance and learning experience effectively by making students watch

online lectures, collaborate in online discussions, or carry out research at home and engage in concepts in the classroom which provides a reason for learning content.

11. Research based learning:
The primary purpose of educational research is to expand the existing body of knowledge by providing solutions to different problems in pedagogy while improving teaching and learning practices. Educational researchers also seek answers to questions bothering on learner-motivation and development.

Assessment Methods / Evaluation Scheme

The assessment of students' achievement in MSc. Big Data Analytics will be aligned with course/program learning outcomes and the academic and technical skills that the program is designed to be developed. A variety of assessment methods that are relevant within the disciplinary area of Data Analytics will be used.

Learning outcomes will be assessed through Students' performance in courses in various ways viz.,

- The oral and written scheduled or weekly tests
- Problem-solving exercises
- Closed-book and open-book tests
- Observing practical skills, practical assignments and documentation
- Individual and group project reports
- Seminar presentations
- Group discussions
- Viva voce examinations
- Digital learning tools
- Literature surveys and evaluations, peers and self-assessment can be the additional methods used.
- Communicating (One and two-way communication; communication within a group, verbal, written and non-verbal communication. Arguing, describing, advocating, interviewing, negotiating, presenting; using specific written forms) by written presentation (essay, report, reflective paper etc.), oral presentation, group work and discussion/debate/role play.
- Regular reading habits in the students need to be inculcated through continuous monitoring and observation about weaker aspect of the Students through peer mentoring.

Discipline Specific– Major Mandatory Courses

Course Code: JPS-**BDA-501-MM**	**Course Title: Probability & Stochastic Process**	**Credits: 3** **Lectures/Week: 3**
Course description	Understand the role of probability theory as well as the concept of random variables and stochastic processes in information and communication technology.	
Learning objectives	● Understand what probability is as well as associated concepts. ● Use probability notations and Stochastic Process	
Course Outcomes	● Know the principle definitions, fundamental theorems, and important relationships in statistics. ● Compute probabilities of transition between states and return to the initial state after long time intervals in Markov chains. ● Identify classes of states in Markov chains and characterize the classes. ● Derive differential equations for time continuous Markov processes with a discrete state space.	
	THEORY	**Total Lectures: 45**
Sub Unit	**Unit – I:**	**45**
1.	a) **Basic Probability**: Concepts of experiments, Outcomes, Sample space, Events, Combinatorial probability, Birthday paradox, Principle of inclusion & exclusion, Conditional probability, Independence, Bayes Theorem. **b) Probability Distribution:** Random Variables: discrete and continuous probability models, some probability distributions: Binomial, Poisson, Geometric, Hypergeometric, Normal, exponential, Chi-square, expectation, variance and other properties of the distribution. **c) Stochastic Process:** Markov Chains, Classification of states, Stationery distribution, limit theorems, Poisson process, illustrations and applications. **d) Introduction to Time Series:**	

	Components of time series, Smoothing auto correlation, stationarity, concepts of AR, MA, ARMA & ARIMA models with illustrations.	
	Evaluation Scheme **I. Continuous Assessment (Internals) - 25 Marks** (i) Internal 1: Test – 10 Marks (ii) Internal 2: Test – 10 Marks (iii)Internal 3: Test – 5 Marks **II. Semester End Examination (SEE)- 50 Marks** Q.1 Answer any two -12 Marks Q.2 Answer any two -12 Marks Q.3 Answer any two -12 Marks Q.4 Answer any two -14 Marks	
References:	1. A First Course in Probability: Shelden M. Ross, 2014. 2. Introduction to Stochastic Process : Paul G. Hoel, Sydney C. Port & Charles J. Stone, Waveland Press, 1987. 3. Time Series Analysis and Its Applications: Robert H. Shumway and David S. Stoffer, Springer 2010.	

Bloom's Taxonomy in Evaluation Scheme

UNIT	KNOWLEDGE	UNDERSTANDING	APPLICATION	TOTAL MARKS
I	20	20	10	50
TOTAL MARKS PER OBJECTIVE	20	20	10	50
% WEIGHTAGE	40	40	20	100

Course Code: **JPS-BDA-501-MMP**	**Practical Title: Probability & Stochastic Process Practical (Credits: 01 Practical/Week: 02 hours)**
	1. Basic Probability 2. Probability Distribution 3. Stochastic Process 4. Introduction to Time Series
	Evaluation Scheme – 25 marks i.e. (50/2 marks) (i) Program – 40 marks (ii) Viva – 5 marks (iii) Journal- 5 marks

Course Code: JPS-BDA-502-MM	**Course Title: Linear Algebra & Linear Programming**	**Credits: 3** **Lectures/Week: 3**
Course description	Use knowledge of content and mathematical procedures/numerical techniques to solve problems and make connections between the different areas of mathematics.	
Learning objectives	● To teach students how to use linear algebra as a powerful tool for computation. ● The second goal is to show how these computations can be conceptualized in a geometric framework. ● The final goal is to give a gentle introduction to the theory of abstract vector spaces.	
Course Outcomes	● Communicate mathematical ideas orally and in writing, with precision, clarity and organization, using proper terminology and notation. ● Demonstrate a solid understanding of rigorous mathematical proof. Students will be able to write clear well-organized and logical mathematical arguments. ● Understanding the Concept of optimization and classification of optimization problems. ● Formulate the LPP for real life Problems and give the solution for the problem using Graphical, Simplex and Big-M method.	

	THEORY	Total Lectures: 45
Sub Unit	**Unit – I:**	**45**
1.	**a) Linear Algebra:** Linear equations and matrices, matrix operations, solving system of linear equations, Gauss-Jordan method, Concept & Computation of determinant and inverse of matrix, Eigen values and eigen vectors, Illustrations of the methods, Positive semi definite and position definite matrices, illustrations. Lab – using R programming **b) Linear Programming:** Definition of the problem, convex sets, corner points, feasibility, basic feasible solutions, Simplex method	
	Evaluation Scheme **I. Continuous Assessment (Internals) - 25 Marks** (i) Internal 1: Test – 10 Marks (ii) Internal 2: Test – 10 Marks (iii)Internal 3: Test – 5 Marks **II. Semester End Examination (SEE)- 50 Marks** Q.1 Answer any two -12 Marks Q.2 Answer any two -12 Marks Q.3 Answer any two -12 Marks Q.4 Answer any two -14 Marks	
References:	1. Linear Algebra and Its Application: Gilbert Strang, 4th Edition, Academic Press. Hands-On Matrix Algebra Using R (Active and Motivated Learning with Applications), Hrishikesh D Vinod, World Scientific 2. Linear Programming: G. Hadley, Addison-Wesley.	

Bloom's Taxonomy in Evaluation Scheme

UNIT	KNOWLEDGE	UNDERSTANDING	APPLICATION	TOTAL MARKS
I	**15**	**20**	**15**	**50**
TOTAL MARKS PER OBJECTIVE	**15**	**20**	**15**	**50**
% WEIGHTAGE	**30**	**40**	**30**	**100**

Course Code: JPS-BDA-502-MMP	**Practical Title: Linear Algebra & Linear Programming Practical (Credits: 01 Practical/Week: 02 hours)**
	1. Linear Algebra 2. Linear Programming
	Evaluation Scheme – 25 marks i.e. (50/2 marks) (i) Program – 40 marks (ii) Viva – 5 marks (iii) Journal- 5 marks

Course Code: JPS-BDA-503-MM	**Course Title: Computing for Data Sciences**	**Credits: 3** **Lectures/Week: 3**
Course description	Understanding computing methodologies for data analysis using R and Python	

Learning objectives	• To demonstrate proficiency with statistical analysis of data. • To apply data science concepts and methods to solve problems in real-world contexts	
Course Outcomes	• Be proficient in using libraries for data analysis in R and Python. • Understand how to use data visualization. • Understand simple statistical summaries using software designed for statistical analyses.	
	THEORY	**Total Lectures: 45**
Sub Unit	**Unit – I:**	**45**
1.	**a) Computer Packages– R and Python:** Usage of R and Python – data handling, data analysis, statistical modeling with illustration in python and R. **b) Data Structure & Concepts of Computation using Java:** Algorithms, Convergence, Complexity with illustrations, some sorting & searching algorithms, some numerical methods e.g. Newton-Raphson, Steepest ascent using Java **c) Computing Methodologies:** Monte-Carlo simulations of random numbers and various statistical methods, memory handling strategies for big data	
	Evaluation Scheme **I. Continuous Assessment (Internals) - 25 Marks** (i) Internal 1: Test – 10 Marks (ii) Internal 2: Test – 10 Marks (iii)Internal 3: Test – 5 Marks **II. Semester End Examination (SEE)- 50 Marks** Q.1 Answer any two -12 Marks Q.2 Answer any two -12 Marks Q.3 Answer any two -12 Marks Q.4 Answer any two -14 Marks	
References:	1. Introduction to Data Science (Data Analysis and Prediction Algorithms with	

	R), Rafael A. Irizarry, https://rafaLabgithub.io/dsbook/ 2. Hands-On Programming with R - Write Your Own Functions and Simulations, Grolemund Garrett, O'Reilly 3. Data Structures and Algorithm using Java, 6th Ed. Michael T. Goodrich and Roberto Tamassia, John Wiley & Sons, Inc 4. Python Data Science Handbook - Essential Tools for Working with Data, Jake VanderPlas, O'Reilly 5. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, WES MCKINNEY, O'Reilly

Bloom's Taxonomy in Evaluation Scheme

UNIT	KNOWLEDGE	UNDERSTANDING	APPLICATION	TOTAL MARKS
I	**15**	**10**	**25**	**50**
TOTAL MARKS PER OBJECTIVE	**15**	**10**	**25**	**50**
% WEIGHTAGE	**30**	**20**	**50**	**100**

Course Code: **JPS-BDA-503-**	**Practical Title: Computing for Data Sciences Practical (Credits: 01 Practical/Week: 02 hours)**

MMP	1. Computer packages :R and Python 2. Data Structure & Concepts of Computation using Java 3. Computing Methodologies
	Evaluation Scheme – 25 marks i.e. (50/2 marks) (i) Program – 40 marks (ii) Viva – 5 marks (iii) Journal- 5 marks

Course Code JPS-BDA-504-MM	**Course Title: MOOC course**	**Credits: 2 Lectures/Week: NA**
Course description		
Learning objectives	● To narrow the gap among students who have been left out of the digital revolution and cannot participate in the knowledge economy. ● to engage students who self-organize their participation based on learning objectives, prior knowledge and abilities, and shared interests	
Course Outcomes	● Helping in the development of the society by providing quality education to every citizen of the country. ● To provide high standard education to the Indian youth learners and strengthen their skills.	
	THEORY	**Total Lectures: NA**
Sub Unit	**Unit – I:**	**NA**
1.	NPTEL SWAYAM course	
	Evaluation Scheme **Grade based students' score**	

Course Code: JPS-**BDA-551-MM**	**Course Title: Enabling Technologies for Data Science-I**	**Credits: 3** **Lectures/Week: 3**
Course description	The course covers importance of data warehousing and different techniques used for data mining and extracting useful information from it.	
Learning objectives	● Be familiar with mathematical foundations of data mining tools. ● Understand and implement classical models and algorithms in data warehouses and data mining ● Characterize the kinds of patterns that can be discovered by association rule mining, classification and clustering. ● Finding knowledge discovery from data warehouse.	
Course Outcomes	● Understand the functionality of the various data mining and data warehousing component ● Appreciate the strengths and limitations of various data mining and data warehousing models ● Explain the analyzing techniques of various data ● Describe different methodologies used in data mining and data warehousing.	
	THEORY	**Total Lectures: 45**
Sub Unit	**Unit – I:**	**45**
1.	**DATA MINING:** **Introduction:** Knowledge discovery from databases, scalability issues. **Data Warehousing:** General principles, modeling, design, implementation and optimization, Cloud Computing, OLAP. **Data Preparation:** Pre-processing, sub-sampling, feature selection. **Classification and Prediction:** Bayes learning, decision trees, CART, neural learning, support vector machines, associations, dependence analysis, rule generation. **Cluster Analysis and Deviation Detection** Partitioning algorithms, Density bases algorithm, Grid based algorithm, Graph theoretic clustering.	

	Temporal and spatial data mining	
	Evaluation Scheme **I. Continuous Assessment (Internals) - 25 Marks** (i) Internal 1: Test – 10 Marks (ii) Internal 2: Test – 10 Marks (iii)Internal 3: Test – 5 Marks **II. Semester End Examination (SEE)- 50 Marks** Q.1 Answer any two -12 Marks Q.2 Answer any two -12 Marks Q.3 Answer any two -12 Marks Q.4 Answer any two -14 Marks	
References:	1. Data Mining Techniques: A. K. Pujari, Sangam Books Ltd., 2001 2. Mastering Data Mining: M. Berry and G. Linoff, John Wiley & Sons., 2000 3. Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management, Olivia Parr Rud, Wiley	

Bloom's Taxonomy in Evaluation Scheme

UNIT	KNOWLEDGE	UNDERSTANDING	APPLICATION	TOTAL MARKS
I	**20**	**10**	**20**	**50**

TOTAL MARKS PER OBJECTIVE	**20**	**10**	**20**	**50**
% WEIGHTAGE	**40**	**20**	**40**	**100**

Course Code: **JPS-BDA-551-MMP**	**Practical Title: Enabling Technologies for Data Science-I Practical (Credits: 01 Practical/Week: 02 hours)**
	1. Data Warehousing: General principles, modeling, design, implementation and optimization, Cloud Computing, OLAP. **2. Data Preparation:** Pre-processing, sub-sampling, feature selection. **3. Classification and Prediction:** Bayes learning, decision trees, CART, neural learning, support vector machines, associations, dependence analysis, rule generation. **4. Cluster Analysis and Deviation Detection:** Partitioning algorithms, Density bases algorithm, Grid based algorithm, Graph theoretic clustering. **5. Temporal and spatial data mining.**
	Evaluation Scheme – 25 marks i.e. (50/2 marks) (i) Program – 40 marks (ii) Viva – 5 marks (iii) Journal- 5 marks

Course Code BDA-552-MM	**Course Title: Foundations of Data Science**	**Credits: 3** **Lectures/Week: 3**
Course description	The course aims to provide fundamental knowledge on graph theory and other important aspects of graph theory.	

Learning objectives	• Able to define the basic concepts of graphs, directed graphs, and weighted graphs. • To understand and apply the fundamental concepts in graph theory • To apply graph theory based tools in solving practical problems • To improve the proof writing skills	
Course Outcomes	• Comprehend fundamental concepts in Data Science and Analytics. • Understand fundamentals of Graph data and apply them to Relational Databases. • Read/Write data of various formats using R. • Understand the concept of dimensions reduction in big data	
	THEORY	**Total Lectures: 45**
Sub Unit	**Unit – I:**	**45**
1.	**a) Graph Theory:** Basic Concepts, Algorithms for connectedness, shortest path, Minimum Sampling Tree, Lab: Graph Databases, Java/Python Programming **b) High Dimensional Space:** Properties, Law of large numbers, Sphere and cube in high dimension, Generating points on the surface of a sphere, Gaussians in High dimension, Random projection, Applications. Lab: Graph Databases, Java/Python Programming **c) Random Graphs:** Large graphs, G(n,p) model, Giant Component, Connectivity, Cycles, Non- Uniform models, Applications. Lab: Graph Databases, Java/Python Programming **d) Singular Value Decomposition (SVD):** Best rank k approximation, Power method for computing the SVD, Applications. Lab: R and Python Programming (Optional: Matlab/Octave) **e) Random Walks**:	

	Reflection Principle, Long leads, Changes of Sign, Illustrations. Lab: R and Python Programming (Optional: Matlab/Octave) **f) Algorithm for Massive Data Problems:** Frequency Moments of data streams, matrix algorithms. Lab: R and Python Programming (Optional: Spark, Matlab/Octave)	
	Evaluation Scheme **I. Continuous Assessment (Internals) - 25 Marks** (i) Internal 1: Test – 10 Marks (ii) Internal 2: Test – 10 Marks (iii)Internal 3: Test – 5 Marks **II. Semester End Examination (SEE)- 50 Marks** Q.1 Answer any two -12 Marks Q.2 Answer any two -12 Marks Q.3 Answer any two -12 Marks Q.4 Answer any two -14 Marks	
References:	1. Foundations of Data Science: John Hopcroft & Ravindran Kannan.	

Bloom's Taxonomy in Evaluation Scheme

UNIT	KNOWLEDGE	UNDERSTANDING	APPLICATION	TOTAL MARKS
I	**20**	**15**	**15**	**50**

TOTAL MARKS PER OBJECTIVE	**20**	**15**	**15**	**50**
% WEIGHTAGE	**40**	**30**	**30**	**100**

Course Code: **JPS-BDA-552-MMP**	**Practical Title: Foundations of Data Science Practical (Credits: 01 Practical/Week: 02 hours)**
	a) Graph Theory: Basic Concepts, Algorithms for connectedness, Shortest path, Minimum Sampling Tree, Lab: Graph Databases, Java/Python Programming **b) High Dimensional Space:** Properties, Law of large numbers, Sphere and cube in high dimension, Generating points on the surface of a sphere, Gaussians in High dimension, Random projection, Applications. Lab: Graph Databases, Java/Python Programming **c) Random Graphs :** Large graphs, G(n,p) model, Giant Component, Connectivity, Cycles, Non-Uniform models, Applications. Lab: Graph Databases, Java/Python Programming **d) Singular Value Decomposition (SVD):** Best rank k approximation, Power method for computing the SVD, Applications. Lab: R and Python Programming (Optional: Matlab/Octave) **e) Random Walks:** Reflection Principle, Long leads, Changes of Sign, Illustrations. Lab: R and Python Programming **f) Algorithm for Massive Data Problems:** Frequency Moments of data streams, matrix algorithms. Lab: R and Python Programming

	Evaluation Scheme – 25 marks i.e. (50/2 marks) (i) Program – 40 marks (ii) Viva – 5 marks (iii) Journal- 5 marks

Course Code JPS-BDA-553-MM	**Course Title: Advanced Statistical Methods**	**Credits: 3** **Lectures/Week: 3**
Course description	The course covers statistical concepts of finding estimation, and formulating hypothesis.	
Learning objectives	● To describe the rationale behind the formulation and components of a statistical model. ● To compare and contrast statistical models in the context of a particular scientific question. ● To communicate statistical ideas to a diverse audience. ● To learn different prediction models.	
Course Outcomes	● Advance statistical concepts and some of their basic applications in real world, ● The appropriate statistical analysis technique for a business problem, ● The appropriateness of statistical analyses, results, and inferences , and, advance data analysis in R ● Interpret the findings from the data analysis, and the implications of those finding**.**	
	THEORY	**Total Lectures: 45**
Sub Unit	**Unit – I:**	**45**
1.	**a) Estimation:** Unbiasedness, Consistency, UMVUE, Maximum likelihood estimates. **b) Test of Hypotheses**: Two types of errors, test statistic, parametric tests for equality of means & variances. **c) Linear Model:** Gauss Markov Model, least square estimators, Analysis of	

	variance. **d) Regression**: Multiple linear regression, forward, backward & stepwise regression, Logistic Regression	
	Evaluation Scheme **I. Continuous Assessment (Internals) - 25 Marks** (i) Internal 1: Test – 10 Marks (ii) Internal 2: Test – 10 Marks (iii)Internal 3: Test – 5 Marks **II. Semester End Examination (SEE)- 50 Marks** Q.1 Answer any two -12 Marks Q.2 Answer any two -12 Marks Q.3 Answer any two -12 Marks Q.4 Answer any two -14 Marks	
References:	1. Statistical Inference: P. J. Bickel and K. A. Docksum, 2nd Edition, Prentice Hall. 2. Introduction to Linear Regression Analysis: Douglas C. Montgomery	

Bloom's Taxonomy in Evaluation Scheme

UNIT	KNOWLEDGE	UNDERSTANDING	APPLICATION	TOTAL MARKS
I	**20**	**20**	**10**	**50**
TOTAL MARKS PER OBJECTIVE	**20**	**20**	**10**	**50**

% WEIGHTAGE	**40**	**40**	**20**	**100**

Course Code: **JPS-BDA-553-MMP**	**Practical Title: Advanced Statistical Methods Practical (Credits: 01 Practical/Week: 02 hours)**
	a) Estimation: Unbiasedness, Consistency, UMVUE, Maximum likelihood estimates. **b) Test of Hypotheses:** Two types of errors, test statistic, parametric tests for equality of means & variances. **c) Linear Model:** Gauss Markov Model, least square estimators, Analysis of variance. **d) Regression:** Multiple linear regressions, forward, backward & stepwiseregression, LogisticRegression
	Evaluation Scheme – 25 marks i.e. (50/2 marks) (i) Program – 40 marks (ii) Viva – 5 marks (iii) Journal- 5 marks

Course Code JPS-BDA-554-MM	**Course Title: Value Thinking**	**Credits: 2** **Lectures/Week: 2**
Course description	**The course is aim to provide or train students to analyze and critical thinking.**	
Learning objectives	● Understand key concepts of critical thinking. ● Clarify the difference in cognition, reasoning and logics. ● Improve their decision making based on facts, assumptions, arguments etc. ● Able to see a problem with a logical approach to find a quick solution.	
Course Outcomes	● Engage the imagination to explore new possibilities. ● Recognize explicit and tacit assumptions and their consequences. ● Distinguish relevant from non-relevant data, fact from opinion. ● Identify, evaluate and synthesize information (obtained through library, world-wide web, and other sources as appropriate) in a	

	collaborative environment.	
	THEORY	**Total Lectures: 30**
Sub Unit	**Unit – I:**	**30**
1.	This course involves watching few movies (list provided below) and reading few books (list provided below) that deals mostly with argumentative logic, evidence, drawing inference from evidences. After watching the movies and reading the books, there will be general discussion amongst the students. Couple of case studies that involve mostly logical thinking will also be presented. Each student will prepare a term paper. Evaluation will be on the basis of this term paper and participation in group discussion. **Movies:** 1. Twelve Angry Men 2. Roshoman by Kurosawa 3. Trial of Nuremberg 4. Mahabharata by Peter Brook **Books:** 1. The Hound of the Baskervilles by Arthur Conan Doyle 2. Five Little Pigs by Agatha Christie 3. The Purloined Letter by Edger Allan Poe 4. The Case of the Substitute Face	
	Evaluation Scheme **I. Continuous Assessment (Internals) - 50 Marks** (i) Internal 1: Test – 10 Marks (ii) Internal 2: Test – 10 Marks (iii)Internal 3: Test – 10 Marks (iv)Internal 4: Test – 10 Marks (v)Internal 5: Test – 10 Marks	

Course Code: JPS-BDA-601-MM	**Course Title: Enabling Technologies for Data Science-II**	**Credits: 3** **Lectures/Week: 3**
Course description	Understand and use the Technologies of Scala, Spark and PySpark for handling data	
Learning objectives	● Understand the basics of functional programming in Scala ● Become comfortable writing Scala code through hands-on labs ● Understand the basics of spark ● Become comfortable with pyspark through hands-on experience	
Course Outcomes	● Be proficient in using libraries for data analysis in Spark and PySpark ● Understand what a machine learning model is and how to use a given model. ● Apply data science concepts and methods to solve problems in real-world contexts and communicate these solutions effectively ● Understand research concepts of using spark and pyspark for data analysis.	
	THEORY	**Total Lectures: 45**
Sub Unit	**Unit – I:**	**45**
1.	Spark, Scala.	
	Evaluation Scheme **I. Continuous Assessment (Internals) - 25 Marks** (i) Internal 1: Test – 10 Marks (ii) Internal 2: Test – 10 Marks (iii)Internal 3: Test – 5 Marks **II. Semester End Examination (SEE)- 50 Marks** Q.1 Answer any two -12 Marks Q.2 Answer any two -12 Marks Q.3 Answer any two -12 Marks Q.4 Answer any two -14 Marks	
References:	NONE	

Bloom's Taxonomy in Evaluation Scheme

UNIT	KNOWLEDGE	UNDERSTANDING	APPLICATION	TOTAL MARKS
I	**10**	**10**	**30**	**50**
TOTAL MARKS PER OBJECTIVE	**10**	**10**	**30**	**50**
% WEIGHTAGE	**20**	**20**	**60**	**100**

Course Code: **JPS-BDA-601-MMP**	**Practical Title: Enabling Technologies for Data Science-II Practical (Credits: 01 Practical/Week: 02 hours)**
	Practical based on theory concepts
	Evaluation Scheme – 25 marks i.e. (50/2 marks) (i) Program – 40 marks (ii) Viva – 5 marks (iii) Journal- 5 marks

Course Code JPS-BDA-602-MM	Course Title: Machine Learning-II	Credits: 3 Lectures/Week: 3
Course description	To learn machine learning and deep learning concepts and their suitability to different scenarios.	
Learning objectives	● To understand the basic theory underlying machine learning. ● To be able to formulate machine learning problems corresponding to different applications. ● To understand a range of machine learning algorithms along with their strengths and weaknesses. ● To be able to apply machine learning algorithms to solve problems of moderate complexity.	
Course Outcomes	● Students learn the different possibilities to deploy large scale machine learning. ● Students explore the different ensemble learning techniques. ● By exploring and implementing case studies students become proficient with the libraries and API to deploy for different Machine learning and deep learning algorithms. ● Students also understand several preprocessing, cross validation and evaluation techniques for different machine learning algorithms	
	THEORY	**Total Lectures: 45**
Sub Unit	**Unit – I:**	**45**
1.	**Decision Tree Classification:** Entropy, Gini index, Algorithms, Regression Trees. **Probabilistic Classifiers:** Generative and Conditional classifiers. **Hyper plane classifiers:** Loss functions, stochastic gradient algorithms, Perceptron algorithms. **Application of to Pattern Recognition Problems.** **Clustering:** Performance criteria, K-means clustering, EM algorithm a) Collaborative filtering	

	b) Combining models c) Probabilistic graphical models d) Large Scale Machine Learning: Gradient descent with large data sets e) Genetic Algorithm.	
	Evaluation Scheme I. Continuous Assessment (Internals) - 25 Marks (i) Internal 1: Test – 10 Marks (ii) Internal 2: Test – 10 Marks (iii)Internal 3: Test – 5 Marks II. Semester End Examination (SEE)- 50 Marks Q.1 Answer any two -12 Marks Q.2 Answer any two -12 Marks Q.3 Answer any two -12 Marks Q.4 Answer any two -14 Marks	
References:	1. Machine Learning: Tom Mitchell	

Bloom's Taxonomy in Evaluation Scheme

UNIT	KNOWLEDGE	UNDERSTANDING	APPLICATION	TOTAL MARKS
I	**15**	**10**	**25**	**50**
TOTAL MARKS PER OBJECTIVE	**15**	**10**	**25**	**50**
% WEIGHTAGE	**30**	**20**	**50**	**100**

Course Code: **JPS-BDA-602-MMP**	**Practical Title: Machine Learning-II Practical (Credits: 01 Practical/Week: 02 hours)**
	Practical based on theory concepts
	Evaluation Scheme – 25 marks i.e. (50/2 marks) (i) Program – 40 marks (ii) Viva – 5 marks (iii) Journal- 5 marks

Course Code JPS-BDA-603-MM	**Course Title: Exploratory Data Analysis**	**Credits: 3 Lectures/Week: 3**
Course description	To learn new ways to collect data, and enable hypothesis testing through experiments and visualization.	
Learning objectives	● Identifying trends in time and space ● Uncover patterns related to the target ● Creating hypotheses and testing them through experiments ● Identifying new sources of data	
Course Outcomes	● Identifying data outliers. ● Detecting patterns of interest, generating hypotheses, ● Creative visualization for decision making data ● Embedding programming scripts to implement the interactive visualization	
	THEORY	**Total Lectures: 45**
Sub Unit	**Unit – I:**	**45**
1.	**Data Visualization with Tableau** Learn about design principles, human perception and effective	

	story telling with data, dashboards, modern visualization tools and techniques (cover Tableau). Hands-on practice on Tableau is must. **Modelling in Operations Management** a) Banking analytics b) Healthcare analytics c) Retail analytics d) Venture analytics e) Marketing analytics f) Supply chain analytics	
	Evaluation Scheme I. Continuous Assessment (Internals) - 25 Marks (i) Internal 1: Test – 10 Marks (ii) Internal 2: Test – 10 Marks (iii)Internal 3: Test – 5 Marks II. Semester End Examination (SEE)- 50 Marks Q.1 Answer any two -12 Marks Q.2 Answer any two -12 Marks Q.3 Answer any two -12 Marks Q.4 Answer any two -14 Marks	
References:	None	

Bloom's Taxonomy in Evaluation Scheme

UNIT	KNOWLEDGE	UNDERSTANDING	APPLICATION	TOTAL MARKS
I	**15**	**10**	**25**	**50**
TOTAL MARKS PER OBJECTIVE	**15**	**10**	**25**	**50**
% WEIGHTAGE	**30**	**20**	**50**	**100**

Course Code: **JPS-BDA-603-MMP**	**Practical Title: Exploratory Data Analysis Practical (Credits: 01 Practical/Week: 02 hours)**
	Practical based on theory concepts
	Evaluation Scheme – 25 marks i.e. (50/2 marks) (i) Program – 40 marks (ii) Viva – 5 marks (iii) Journal- 5 marks

Discipline Specific– Major Elective Courses

Course Code JPS-BDA-505-ME	**Course Title: Database Management**	**Credits:3 Lectures/Week: 3**
Course description	Understand the basic concepts and the applications of database systems and building blocks of Big Data	
Learning objectives	• Introduction to database management systems, with an emphasis on how to organize, maintain and retrieve - efficiently, and effectively - information from a DBMS. • To understand Big Data and its applications	
Course Outcomes	• Able to perform the basic elements of a relational database management system. • Identify the data models for relevant problems. • To understand the applications using Map Reduce Concepts. • To introduce programming tools PIG & HIVE in Hadoop ecosystem.	
	THEORY	**Total Lectures: 45**
Sub Unit	**Unit – I:**	**45**
1.	**a) Basic Concepts:** Different data models, ER and EER diagram, schema, table, Big Data Concepts and Hadoop Ecosystem **b) Relational and Non-Relational Databases:** Structure, various operations, normalization, SQL, No-SQL, Graph Database, Parallel and distributed data base, Map-Reduce. Lab using SQL/Oracle/MySql for Relational databases; Hadoop(any), MangoDB, GraphDB for Big Data **c) Implementation:** ORACLE SQL/MS SQL/MySQL, Hadoop Ecosystem, Concept	

	of database security	
	Evaluation Scheme **I. Continuous Assessment (Internals) - 25 Marks** (i) Internal 1: Test – 10 Marks (ii) Internal 2: Test – 10 Marks (iii)Internal 3: Test – 5 Marks **II. Semester End Examination (SEE)- 50 Marks** Q.1 Answer any two -12 Marks Q.2 Answer any two -12 Marks Q.3 Answer any two -12 Marks Q.4 Answer any two -14 Marks	
References:	1. Database system concepts: Abraham Silberschartz, Henry F. Korth and S. Surarshan, McGraw Hill, 2011. 2. Hadoop 2 Quick-Start Guide: Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem, Douglas Eadline, Addison-Wesley, Pearson Education India; First edition (1 March 2016) 3. Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, EMC Education Services, 2015	

Bloom's Taxonomy in Evaluation Scheme

UNIT	KNOWLEDGE	UNDERSTANDING	APPLICATION	TOTAL MARKS
I	**10**	**15**	**25**	**50**
TOTAL MARKS PER OBJECTIVE	**10**	**15**	**25**	**50**

% WEIGHTAGE	**20**	**30**	**50**	**100**

Course Code: **JPS-BDA-505-MEP**	**Practical Title: Database Management Practical (Credits: 01 Practical/Week: 02 hours)**
	1. Relational Databases using MS SQL 2. Non-relational databases 3. Implementations
	Evaluation Scheme – 25 marks i.e. (50/2 marks) (i) Program – 40 marks (ii) Viva – 5 marks (iii) Journal- 5 marks

Course Code: JPS-BDA-555-ME	**Course Title: Cloud Computing**	**Credits:3 Lectures/Week: 3**
Course description	**The course delivers the hands on experience to understand and use different cloud services.**	
Learning objectives	• This course gives students an insight into the basics of cloud computing along with virtualization • It will provide the students basic understanding about cloud and virtualization along with it how one can migrate over it. • Understand the concept of Cloud Security. • Learn the Concept of Cloud Infrastructure Model.	
Course Outcomes	• Able to store, manage, process, share, collaborate data and information with high speed and accuracy. • Understand Platform as a Service, Infrastructure as a service and Software as Service • Understand the new ways you can use to program, develop, deploy and provide application access to the users	

	• Able to identify problems, and explain, analyze, and evaluate various cloud computing solutions.	
	THEORY	**Total Lectures: 45**
Sub Unit	**Unit – I:**	**45**
1.	Introduction to Cloud computing, Cloud service methods, IaaS, PaaS, SaaS, fundamentals of cloud Architecture (load distribution, resource pooling, scalability, load balancing, redundancy, etc), Introduce DevOps, CICD.	
	Evaluation Scheme **I. Continuous Assessment (Internals) - 25 Marks** (i) Internal 1: Test – 10 Marks (ii) Internal 2: Test – 10 Marks (iii)Internal 3: Test – 5 Marks **II. Semester End Examination (SEE)- 50 Marks** Q.1 Answer any two -12 Marks Q.2 Answer any two -12 Marks Q.3 Answer any two -12 Marks Q.4 Answer any two -14 Marks	

Bloom's Taxonomy in Evaluation Scheme

UNIT	KNOWLEDGE	UNDERSTANDING	APPLICATION	TOTAL MARKS
I	**10**	**10**	**30**	**50**
TOTAL MARKS PER OBJECTIVE	**10**	**10**	**30**	**50**

% WEIGHTAGE	**20**	**20**	**60**	**100**

Course Code: JPS-BDA-555-MEP	**Practical Title: Cloud Computing Practical (Credits: 01 Practical/Week: 02 hours)**
	Hands-on practice either on AWS/Azure/Google Cloud Platform
	Evaluation Scheme – 25 marks i.e. (50/2 marks) (i) Program – 40 marks (ii) Viva – 5 marks (iii) Journal- 5 marks

Course Code JPS-BDA-604-ME	**Course Title: Introduction to Econometrics and Finance**	**Credits:3 Lectures/Week: 3**
Course description	To understand the social science studies and field-econometrics used in Big data analytics	
Learning objectives	● To introduce regression analysis to students so that they are able to understand its applications in different fields in economics. ● to establishing trends between datasets ● to get a specific pattern or result from cluttered data. ● to give necessary tools to conduct empirical research.	
Course Outcomes	● Optimize portfolio on the collected historical Sensex data of different company for giving maximum return with minimum risk. ● Analyze the pattern of return of different company from historical Sensex data. ● Predict the return for a certain amount of time for different company and to check their prediction accuracy from the actual data. ● Apply Binomial Model in real life Put Call parity problems and also understand model working procedure by simulated data.	

	• Apply Black Sholes formula in real life scenarios and also on simulated data	
	THEORY	**Total Lectures: 45**
Sub Unit	**Unit – I:**	**45**
1.	a) Analysis of Panel Data. b) Generalized Method of Moments (GMM). c) Simultaneous Equations System: Least Squares, Bias Problem, Estimation Method. d) Cointegration: Concept, two variable model, Engle-Granger Method, Vector auto regressions (VAR), Vector error correlation model (VECM). e) ARCH/GARCH/SV models, some important generalizations like EGARCH & GJR models, ARCH –M models.	
	Evaluation Scheme I. Continuous Assessment (Internals) - 25 Marks (i) Internal 1: Test – 10 Marks (ii) Internal 2: Test – 10 Marks (iii)Internal 3: Test – 5 Marks II. Semester End Examination (SEE)- 50 Marks Q.1 Answer any two -12 Marks Q.2 Answer any two -12 Marks Q.3 Answer any two -12 Marks Q.4 Answer any two -14 Marks	
References:	1. The Econometrics of Financial Markets: J. Campbell, A.Lo and C. Mackinlay 2. Econometric Analysis: William H. Greene	

Bloom's Taxonomy in Evaluation Scheme

UNIT	KNOWLEDGE	UNDERSTANDING	APPLICATION	TOTAL MARKS
I	**15**	**20**	**15**	**50**
TOTAL MARKS PER OBJECTIVE	**15**	**20**	**15**	**50**
% WEIGHTAGE	**30**	**40**	**30**	**100**

Course Code: **JPS-BDA-604-MEP**	**Practical Title: Introduction to Econometrics and Finance Practical (Credits: 01 Practical/Week: 02 hours)**
	Practical based on theory concepts
	Evaluation Scheme – 25 marks i.e. (50/2 marks) (i) Program – 40 marks (ii) Viva – 5 marks (iii) Journal- 5 marks

Research Methodology Courses

Course Code: JPS-BDA-506-RM	**Course Title: Statistical Methods**	**Credits:3 Lectures/Week: 3**
Course description	Understand the fundamental statistical concepts and some of their basic applications in real world.	
Learning objectives	● Defining the type and quantity of data need to be collected. ● Organizing and summarizing the data. Analyzing the data and drawing conclusions from it.	
Course Outcomes	● How to use a wide variety of specific statistical methods, and computer programming in R. ● Apply technologies in organizing different types of data. ● Present results effectively by making appropriate displays, summaries, and tables of data, perform simple statistical analyses using R. ● Analyze the data and come up with correct interpretations and relevant conclusions.	
	THEORY	**Total lectures: 45**
Sub Unit	**Unit – I:**	**45**
1.	**a) Data Collection & Visualization:** Concepts of measurement, scales of measurement, design of data collection formats with illustration, data quality and issues with date collection systems with examples from business, cleaning and treatment of missing data, principles of data visualization, and different methods of presenting data in business analytics. **b) Basic Statistics:**	

	Frequency table, histogram, measures of location, measures of spread, skewness, curtosis, percentiles, box plot, correlation and simple linear regression, partial correlation, probability distribution as a statistics model, fitting probability distributions, empirical distributions, checking goodness of fit through plots and tests. **c) Contingency Tables:** Two way contingency tables, measures of association, testing for dependence.	
	Evaluation Scheme I. Continuous Assessment (Internals) - 100 Marks (i) Internal 1: Test – 20 Marks (ii) Internal 2: Test – 20 Marks (iii)Internal 3: Test – 20 Marks (iv) Internal 4: Test - 20 Marks (v) Internal 5 :Test - 20 Marks	
References:	1. Statistics: David Freedman, Pobert Pisani & Roger Purves, WW.Norten & Co. 4th Edition 2007. 2. The visual display of Quantitative Information: Edward Tufte, Graphics Press, 2001. 3. Best Practices in Data Cleaning: Jason W. Osborne, Sage Publications 2012.	

Bloom's Taxonomy in Evaluation Scheme

UNIT	KNOWLEDGE	UNDERSTANDING	APPLICATION	TOTAL MARKS
I	**30**	**30**	**40**	**100**
TOTAL MARKS PER OBJECTIVE	**30**	**30**	**40**	**100**

% WEIGHTAGE	**30**	**30**	**40**	**100**

Course Code: **JPS-BDA-506-RMP**	**Practical Title: Statistical Methods Practical (Credits: 01 Practical/Week: 02 hours)**
	1.Data Collection & Visualization 2. Basic Statistics 3. Contingency Tables

Field Project Courses

Course Code: JPS-BDA-556-FP	**Course Title: Machine Learning-I**	**Credits:3 Lectures/Week: 3**
Course description	The course provide extensive knowledge on different machine and deep learning concepts.	
Learning objectives	• To understand the basic theory underlying machine learning. • To be able to formulate machine learning problems corresponding to different applications. • To understand a range of machine learning algorithms along with their strengths and weaknesses. • To be able to apply machine learning algorithms to solve problems of moderate complexity.	
Course Outcomes	• pre-process and analyze the characteristics of different types of standard data, • work on scikit-learn, a standard machine learning library, • Evaluate the performance of different machine learning techniques for a particular application and validate the significance of the results obtained. • Build skills to implement different classification and clustering techniques as per requirement to extract valuable information from any type of data set. .	

	THEORY	45 lectures
Sub Unit	**Unit – I:**	**45**
1.	**a. Linear Regression** Linear Regression with Multiple variables, applications. **b. Logistic Regression:** Model, Classification, Problem of over-fitting, Applications. **c. Neural Networks:** Representation Learning, Different Models like single and multi-layer perceptron, back propagation, Application. **d. Machine Learning System Design:** Evaluating a learning algorithms, handling skewed data, using large data sets. **e. Support Vector Machines:** Model, Large Margin Classification, Kernels, SVMs in practice. f. **Unsupervised Learning**. **g. Dimensionality Reduction.** **h. Anomaly Detection**.	
	Evaluation Scheme **I. Continuous Assessment (Internals) - 25 Marks** (i) Internal 1: Test – 10 Marks (ii) Internal 2: Test – 10 Marks (iii)Internal 3: Test – 5 Marks **II. Semester End Examination (SEE)- 50 Marks** Q.1 Answer any two -12 Marks Q.2 Answer any two -12 Marks Q.3 Answer any two -12 Marks Q.4 Answer any two -14 Marks	
References:	Machine Learning: Tom Mit chell	

Bloom's Taxonomy in Evaluation Scheme

UNIT	KNOWLEDGE	UNDERSTANDING	APPLICATION	TOTAL MARKS
I	**20**	**15**	**15**	**50**
TOTAL MARKS PER OBJECTIVE	**20**	**15**	**15**	**50**
% WEIGHTAGE	**40**	**30**	**30**	**100**

Course Code: **JPS-BDA-556-FPP**	**Practical Title: Machine Learning-I Practical (Credits: 01 Practical/Week: 02 hours)**
	a.Linear Regression Linear Regression with Multiple variables, applications. **b.Logistic Regression:** Model, Classification, Problem of over-fitting, Applications. **c.Neural Networks:** Representation Learning, Different Models like single and multi-layer perceptron, back propagation, Application. **d.Machine Learning System Design:** Evaluating a learning algorithms, handling skewed data, using large data sets. **e.Support Vector Machines:** Model, Large Margin Classification, Kernels, SVMs in practice. **f.Unsupervised Learning** **g.Dimensionality Reduction.**

	h.Anomaly Detection
	Evaluation Scheme – 25 marks i.e. (50/2 marks) (i) Program – 40 marks (ii) Viva – 5 marks (iii) Journal- 5 marks

Research Project Courses

Course Code: JPS-BDA-605-RP	**Course Title: Introduction to Bioinformatics**	**Credits:3 Lectures/Week:3**
Course description	Introduction to the basic concepts of Bioinformatics and its significance in Biological data analysis.	
Learning objectives	● Apply reasoning about core biological concepts with emphasis on the cellular and molecular scale of biology ● Design, implement and evaluate computer-based systems, processes, components or programs in relation to the contexts of molecular and cellular biology and genomics research. ● Analyze and evaluate bioinformatics data to discover patterns, critically evaluate conclusions and generate predictions for subsequent experiments. ● Communicate biological information relating to bioinformatics in both written and oral forms	
Course Outcomes	● Describe the history, scope and importance of Bioinformatics and role of internet in Bioinformatics. ● Explain about the methods to characterize and manage the different types of Biological data. ● Introduction to the basics of sequence alignment and analysis. ● Classify different types of Biological Databases.	

	THEORY	Total lectures: 45
Sub Unit	**Unit – I:**	**45**
1.	a) Sequence Alignments. b) Advance Alignment Methods. c) Gibbs Sampling. d) Population Genomics. e) Genetic Mapping. f) Disease Mapping. g) Gene Recognition. h) Transcriptome & Evolution. i) Protein Structure. j) Protein Motifs. k) Hidden Markov Model. l) Lattice Model. m) Algorithms.	
	Evaluation Scheme I. Continuous Assessment (Internals) - 25 Marks (i) Internal 1: Test – 10 Marks (ii) Internal 2: Test – 10 Marks (iii)Internal 3: Test – 5 Marks II. Semester End Examination (SEE)- 50 Marks Q.1 Answer any two -12 Marks Q.2 Answer any two -12 Marks Q.3 Answer any two -12 Marks Q.4 Answer any two -14 Marks III. Research Project - 50 marks (for 2 credits)	
References:	1. Introduction to Computational Molecular Biology: C. Setubal & J. Meidanis, PWS Publishing, Boston, 1997	

Bloom's Taxonomy in Evaluation Scheme

UNIT	KNOWLEDGE	UNDERSTANDING	APPLICATION	TOTAL MARKS
I	**15**	**15**	**20**	**50**
TOTAL MARKS PER OBJECTIVE	**15**	**15**	**20**	**50**
% WEIGHTAGE	**30**	**30**	**40**	**100**

Course Code: **JPS-BDA-605-RPP**	**Practical Title: Introduction to Bioinformatics Practical (Credits: 01 Practical/Week: 02 hours)**
	Practical based on theory concepts

On Job Training

Course Code: JPS-BDA-651-OJT	Course Title: Internship based project	Credits: 22 Lectures/Week: NA
Course description	To work on project in corporate world	
Learning objectives	● Help the student develop written communication skills. ● Serve as an archival record of the internship experience. ● Give the student an opportunity to reflect on the professional aspects of the internship experience and the skills that were learned. ● to build an understanding of the knowledge and skills in an industry or workplace	
Course Outcomes	● Acquire on job the skills, knowledge, and attitude, which are requisite to constitute a professional identity. ● learn to organize data projects and provide work experience that can lead directly to full-time positions ● Knowledge to examine information using data analysis tools so that they can help their employers make important decisions ● student can practice and improve their industry skills while also gaining industrial experience ● Help a student start to build a professional network that can be a resource for the student.	
	THEORY	**NA**
Sub Unit	**Unit – I:**	
1.	A real-life project has to be undertaken at an industry for 20 weeks. Each student will have two supervisors: one from academic institution and one from the industry. The project shall involve handling data extensively and use of methodologies learnt during the course work to derive meaningful inferences. A final project report has to be submitted and an "open" presentation has to be made.	
	Evaluation Scheme ● Report from two supervisors:	

	200 marks (100 each) Project report: 200 marks ● Presentation: 300 marks. ● Total: 500 marks	