HSNC UNIVERSITY, MUMBAI

HSNCU Syllabus

School of Applied Sciences

Syllabus of BSc Data Science and Business Analytics

Board of Faculty of Science & Technology

Board of Studies in the Subject of Data Science & Business Analytics:

1. Name of the Chairperson: Dr. M A Khan, Ex Registrar, University of Mumbai, Ex Director, Sydenham Management Institute, Ex Professor and Director, NL Dalmia Institute of Management studies, Ex CEO, HCOI Ministry of Minority Affairs, GOI

2. Name of the Co – Chairperson: Mrs. Shailaja. J. Rane, Academic head of Data Science and Business Analytics, HSNC University, Mumbai.

Teachers from the University:

3. Miss. Beenarani Karutharan, Assistant Professor, Department of Computer science, KC College, HSNC University, Mumbai.

4. Mrs. Mrunal M Hardikar , Assistant Professor, Department of Mathematics, KC College, HSNC University, Mumbai.

5. Miss. Anjali Sutar, Assistant Professor, Department of Data Science and Business Analytics, HSNC University, Mumbai.

6. Miss. Shweta Maitri, Assistant Professor, Department of Data Science and Business Analytics, HSNC University, Mumbai.

7. Miss. Shruti Agarwal, Assistant Professor, Department of Data Science and Business Analytics, HSNC University, Mumbai.

External Professors:

8. Rosemary Gosling, Director of External Studies(Retd) for the London School of Economics and Political Science (LSE)

9. Dr. Santosh Bothe, Founder and Director AiSense(Start-up funded by BIRAC, Govt. of India), Principal, Saraswati College, Shegaon, Affiliated to SGBU Amravati University.

External experts:

10. Prof. Parag Mahulikar is Ex- Dean and Senior Professor of Marketing at IES Management College and Research Centre, Bandra, India and Management Consultant.

11. Dr. Alok Deepak Dabade, Assistant Professor, Department of Statistics, University of Mumbai.

12. Dr. Sujata Suvarnapathki, Assistant Professor, Department of Statistics, Ramnarayan Ruia Autonomous College, Matunga, Mumbai.

13. Mr. Subhash Kumar, Assistant Professor, MCA, MPHIL IT department, St.Xavier's college, Mumbai.

14. Industry Expert: Mr. Vinayak Deshpande, Managing Director, Sankhya Analytical Research Pvt. Ltd.

15. Industry Expert: Mr. Nishad Kapadia, MCA, Technical trainer,Data and Solution Architect Project Manager, TeraData, Mumbai.

16. Industry Expert: Miss. Praveena Premanand Menon, MSc in Big Data Analytics.

17.Industry Expert: Mr. Awesh Bhornya ,Infinity Learning (Founder)

Alumni: Proposed names: Rushabh Maru, Harsh Mahapadi.

Part –I

Outline of Choice Based Credit System as outlined by University Grants Commission:

The Definitions Of The Key Terms Used In The Choice Based Credit System And Grading System Introduced From The Academic Year 2023-2024 Are As Under:

1. **Core Course:** A course, which should compulsorily be studied by a candidate as a core requirement is termed as a core course.

2. **Elective Course:** Generally, a course which can be chosen from a pool of courses and which may be very specific or specialized or advanced or supportive to the discipline/subject of study or which provides an extended scope or which enables an exposure to some other discipline/subject/domain or nurtures the candidate's proficiency/skill is called an Elective Course.

2.1 **Discipline Specific Elective (DSE) Course:** Elective courses may be offered by the main discipline/subject of study is referred to as Discipline Specific Elective. The University/Institute may also offer discipline related Elective courses of interdisciplinary nature (to be offered by main discipline/subject of study).

2.2 **Dissertation/Project:** An elective course designed to acquire special/advanced knowledge, such as supplement study/support study to a project work, and a candidate studies such a course on his own with an advisory support by a teacher/faculty member is called dissertation/project. A Project/Dissertation work would be of 6 credits. A Project/Dissertation work may be given in lieu of a discipline specific elective paper.

2.3 **Generic Elective (GE) Course:** An elective course chosen generally from an unrelated discipline/subject, with an intention to seek exposure is called a Generic Elective.

P.S.: A core course offered in a discipline/subject may be treated as an elective by other discipline/subject and vice versa and such electives may also be referred to as Generic Elective.

3. **Choice Base Credit System:** CBCS allows students to choose inter- disciplinary, intra-disciplinary courses, skill-oriented papers (even from other disciplines according to their learning needs, interests and aptitude) and more flexibility for students.

4. **Honours Program:** To enhance employability and entrepreneurship abilities among the learners, through aligning Inter Disciplinary / Intra Disciplinary courses with Degree Program. Honours Program will have 40 additional credits to be undertaken by the learner across three years essentially in Inter / Intra Disciplinary course.

A learner who joins Regular Undergraduate Program will have to opt for Honours Program in the first year of the Program. However, the credits for honors, though divided across three years can be completed within three years to become eligible for award of honours Degree.

5. **Program:** A Program is a set of course that are linked together in an academically meaningful way and generally ends with the award of a Degree Certificate depending on the level of knowledge attained and the total duration of study.

6. **Course:** A 'course' is essentially a constituent of a 'program' and may be conceived of as a composite of several learning topics taken from a certain knowledge domain, at a certain level. All the learning topics included in a course must necessarily have academic coherence, i.e. there must be a common thread linking the various components of a course. A number of linked courses considered together are in practice, a 'program'.

7. **Bridge Course:** Bridge course is visualized as Pre-semester preparation by the learner before commencement of regular lectures. For each semester the topics, whose knowledge is considered as essential for effective and seamless learning of topics of the Semester, will be specified. The Bridge Course can be conducted in online mode. The Online content can be created for the Bridge Course Topics.

8. **Module and Unit:** A course which is generally an independent entity having its own separate identity, is also often referred to as a 'Module' in today's parlance, especially when we refer to a 'modular curricular structure'. A module may be studied in conjunction with other learning modules or studied independently. A topic within a course is treated as a Unit. Each course should have exactly 3 Units.

9. **Self-Learning: 20% of the topics will be marked for Self-Learning.** Topics for Self-Learning are to be learned independently by the student, in a time- bound manner, using online and offline resources including online lectures, videos, library, discussion forums, fieldwork, internships etc.

Evaluative sessions (physical/online), equivalent to the credit allocation of the Self Learning topics, shall be conducted, preferably, every week for each course.

Learners are to be evaluated real time during evaluative sessions. The purpose of evaluative sessions is to assess the level of the students' learning achieved in the topics earmarked for Self-Learning.

The teacher's role in these evaluative sessions will be that of a Moderator and Mentor, who will guide and navigate the discussions in the sessions, and offer concluding remarks, with proper reasoning on the aspects which may have been missed by the students, in the course of the Self-Learning process.

The modes to evaluate self-learning can be a combination of the various methods such as written reports, handouts with gaps and MCQs, objective tests, case studies and Peer learning. Groups can be formed to present self-learning topics to peer groups, followed by Question and Answer sessions and open discussion. The marking scheme for Self-Learning will be defined under Examination and Teaching.

The topics stipulated for self-learning can be increased or reduced as per the recommendations of the Board of Studies and Academic Council from time to time. All decisions regarding evaluation need to be taken and communicated to the stakeholders preferably before the commencement of a semester. Some exceptions may be made in exigencies, like the current situation arising from the lockdown, but such ad hoc decisions are to be kept

to the minimum possible.

10. **Credit Point:** Credit Point refers to the 'Workload' of a learner and is an index of the number of learning hours deemed for a certain segment of learning. These learning hours may include a variety of learning activities like reading, reflecting, discussing, attending lectures / counselling sessions, watching especially prepared videos, writing assignments, preparing for examinations, etc. Credits assigned for a single course always pay attention to how many hours it would take for a learner to complete a single course successfully. A single course should have, by and large a course may be assigned anywhere between 2 to 8 credit points wherein 1 credit is construed as corresponding to approximately 15/30 learning hours.

11. **Credit Completion and Credit Accumulation:** Credit completion or Credit acquisition shall be considered to take place after the learner has successfully cleared all the evaluation criteria with respect to a single course. Thus, a learner who successfully completes a 4 CP (Credit Point) course may be considered to have collected or acquired 4 credits. Learner level of performance above the minimum prescribed level (viz. grades / marks obtained) has no bearing on the number of credits collected or acquired. A learner keeps on adding more and more credits as he completes successfully more and more courses. Thus, the learner 'accumulates' course wise credits.

12. **Credit Bank:** A Credit Bank in simple terms refers to stored and dynamically updated information regarding the number of Credits obtained by any given learner along with details regarding the course/s for which Credit has been given, the course-level, nature, etc. In addition, all the information regarding the number of Credits transferred to different programs or credit exemptions given may also be stored with the individual's history.

13. **Credit Transfer:** (performance transfer) When a learner successfully completes a program, he/she is allowed to transfer his/her past performance to another academic program having some common courses and Performance transfer is said to have taken place.

14. **Course Exemption:** Occasionally, when two academic programs offered by a single university or by more than one university, may have some common or equivalent course-content, the learner who has already completed one of these academic programs is allowed to skip these 'equivalent' courses while registering for the new program. The Learner is 'exempted' from 'relearning' the common or equivalent content area and from re-appearing for the concerned examinations. It is thus taken for granted that the learner has already collected in the past the credits corresponding to the exempted courses.

Part-II

The Scheme of Teaching and Examination:

The performance of the learners shall be evaluated in two components: Internal Assessment with 15 marks by way of continuous evaluation and by Semester End Examination with 60 marks by conducting the theory examination.

INTERNAL ASSESSMENT: - It is defined as the assessment of the learners on the basis of continuous evaluation as envisaged in the credit-based system by way of participation of learners in various academic and correlated activities in the given semester of the programme.

A). Formative assessment –

1. For Theory Courses

Sr. No.	Particulars	Marks
1	Self-Learning Evaluation	20 Marks

2. For Courses with Practicals:

I.

Sr. No.	Particulars	Marks
1	Self-Learning Evaluation	15 Marks

II. Practical Examination will be conducted out of 25 marks

Sr. No	Evaluation type	Marks
1	Journal	2.5
2	Viva	2.5
3	Practical exam	20

B) Summative Assessment:

The semester end examination of 60% for each course will be as follows:

i) **Duration** – **2 Hours**

ii) **Theory Question Paper Pattern**: -

1. There shall be four questions each of 15 marks. There will be one question each from each unit and the fourth one will be based on the entire syllabus.

2. All questions shall be compulsory and internal choice within the questions is available. (Each question will be of 25 marks with options.)

3. Question may be subdivided into sub-questions a, b, c, d, e of 5 marks each.

The marks will be given for all examinations and they will be converted into grade (quality) points. The semester-end, final grade sheets and transcripts will have Marks, credits, grades, grade points, SGPA and CGPA.

C) Project:

Project can be in any of the following forms

− Case Studies

− Videos

− Blogs

− Research paper (Presented in Seminar/Conference)

− Field Visit Report

− Presentations related to the subject (Moot Court, Youth Parliament, etc.)

− Internships (Exposition of theory into practice)

− Open Book Test

− any other innovative methods adopted with the prior approval of Director Board of Examination and Evaluation.

Self-Learning Evaluation:

− 20% of the topics of curriculum are learned by the student through self-learning using online/ offline academic resources specified in the curriculum, hence 20% of the lectures shall be allocated for evaluation of students on self-learning topics.

− The identified topics in the syllabus shall be learnt independently by the students in a time bound manner preferably from online resources. Evaluative sessions shall be conducted by the teachers and will carry 15/20 marks.

Earmark self-learning evaluation lectures in the timetable. Evaluation of self-learning topics can commence in regular lectures assigned for self-learning evaluation in the timetable.

Methods for Evaluation of Self-learning topics:

− Seminars/presentation (PPT or poster), followed by Q&A − Objective questions /Quiz / Framing of MCQ questions.
− Debates
− Group discussion
− You-Tube videos (Marks shall be based on the quality and viewership)
− Improvisation of videos
− Role Play followed by question-answers
− Viva Voce
− Any other innovative method

Teachers can frame other methods of evaluation also provided that the method, duly approved by the college examination committee, is notified to the students at least 7 days before the commencement of the evaluation session and is forwarded for information and necessary action at least 3 days before the commencement of the evaluation session.

The marks of the internal assessment should not be disclosed to the students till the results of the corresponding semester is declared.

HSNC University, Mumbai

(2025-2026)

Ordinances and Regulations

With Respect to

Choice Based Credit System (CBCS)

For the Programmes Under

The Faculty of Science and Technology

For the Course

Data Science & Business Analytics

Curriculum – Third Year Undergraduate Program

Semester-V and Semester -VI

Data Science & Business Analytics

Part 1- Preamble

B. Sc. Data Science and Business Analytics program is of minimum 140 credits cover six semesters. Data is the new oil. The analytics may be input for human decisions or may drive fully automated decisions. It helps decision maker in building strategies to perform deep-dive understanding and provide descriptive, predictive, and prescriptive analytics. It is used to run the business effectively and is instrumental in growing the business. It is the area for huge potential for corporate investments. Business Analytics include identifying KPIs, measurement strategy, data analysis, complex statistical model and analysis, data mining and deep understanding of cause-and-effect models. Business analytics can drive key decision making in the organization and help executive decision makers in building strategy, predictive analysis, forecasting, risk analysis, identify and prevent fraud, market analysis, etc. Data Scientists use these skills to provide insights into discrete data sets, build complex model and present them in Scorecard format and use the same in executive reviews to lead data-driven discussion and decisions. Some of the impactful use of this is in the areas of Management Information Systems, Financial Service, Marketing Research, Process Improvements, Six Sigma, Process Excellence, Scorecard, Dashboard, End-to-End Product Management, etc.

The program emphasizes both theory and modern applications of Data Science and Business analytics and is structured to provide knowledge and skills in depth necessary for the employability of students in industry, in academics and other government and non-government organizations. The program has some unique features like independent projects, number of elective courses and extensive computer training of statistical computations including standard software packages like C++, SQL, SPSS, SAS, MINITAB, R and PYTHON etc. Due to Cluster University, the department got the academic autonomy and it has been utilized to add the new and need based elective courses. The independent project work is one among the important components of this program. The syllabus has been framed to possess a decent balance of theory, methods, and applications of statistics. It is possible for the students to study basic courses from other disciplines like economics, life sciences, computer science and Information Technology in place of optional/electives. The thrust of the course is to prepare students to enter a promising career after graduation, as also provide to them a platform for pursuing higher studies resulting in postgraduation degrees.

2. Salient features, how it has been made more relevant.

➢ Syllabus is prepared by top most industry experts, eminent statisticians, Data Scientists and reputed faculties of HSNC University.

➢ Skills covered includes Statistics, Data Analytics, Business Analytics, Data Visualization, machine learning algorithm, Text Mining, Forecasting/ Predictive Analytics, Hypothesis Testing etc.

➢ Tools and Technology covered includes R, Python, My SQL, NO SQL, Tableau, Power BI, Machine Learning, Artificial Intelligence, SPSS, AMOS, Minitab, Excel, and many more.

➢ Course includes several Assignments and Case Studies.

➢ Getting industrial experience by working on industry relevant live projects and internship.

➢ University has a dedicated placement cell for the participants for internships and placements.

➢ Teaching Faculties will include good blend of Academicians and Industry Experts.

➢ Focus is to prepare Participants with clear, concise concept to experts in data Science field to add quality and value to institute they join.

➢University/ College has well placed Alumni working at top positions.

Program Outcomes:

1. **Understanding Core Scientific Concepts**: Students will gain clear insight and understanding to recall key scientific principles across various fields. A well-established foundational knowledge of the subject will play a crucial role for deeper learning and future studies.
2. **Commit to Lifelong Scientific Learning**: Students will cultivate a habit of continuous learning and shall learn to stay updated with the latest scientific and technological advancements. This mindset will ensure that they remain relevant, engaged and informed throughout their future academic journey.
3. **Abilities to Analyse and Evaluate**: Students will learn to classify and scrutinize complex problems into manageable parts, critically analyse data, and evaluate potential solutions to scientific problems.
4. **Assessing Ethical Implications**: Students will be trained to evaluate the ethical dimensions of research and technological innovations, ensuring that their decisions consider societal impacts and they adhere to ethical standards. This is vital for responsible and sustainable practices.
5. **Design Experiments and Innovate**: Students will learn to design and conduct experiments, developing innovative solutions to challenges through Research Projects. They will also learn to evaluate their results and refine their experimental approaches over time.
6. **Application of Scientific and Technical Knowledge to Real-World Problems:** Students will use their scientific and Technical knowledge and expertise to identify and solve real-world problems. This would involve applying theoretical concepts to practical situations, bridging the gap between classroom learning, Industry-academia and real-life applications.

7. **Communicating Scientific Findings Effectively**: Students will develop the ability to communicate scientific information clearly and effectively, both in writing and verbally. Whether presenting research findings or writing technical reports, clear communication is key to knowledge sharing and collaboration.
8. **Foster an Interdisciplinary Approach**: Students will cultivate leadership and teamwork skills, enabling them to collaborate effectively in diverse, interdisciplinary teams. Leadership qualities such as decision-making and delegation will help them achieve successful outcomes in various projects.
9. **Promote Environmental Sustainability**: Students will understand the environmental impact of scientific activities and advocate for sustainable practices. By considering environmental factors in their work, they will contribute to the long-term health of the planet.
10. **Enhanced Skills for Entrepreneurship and Employability**: Students will be equipped with essential skills for entrepreneurship and employability, focusing on job readiness, soft skills, and practical business knowledge. Hands-on experience through internships and mentorship will further enhance their readiness for the job market and entrepreneurial ventures.

Programme Educational Objectives (PEOS)

PEO1. To develop problem-solving skills using data-driven techniques and advanced analytics.

PEO2. To build the necessary skill set and analytical abilities for applying data science and business analytics techniques to real-world problems across various industries.

PEO3. To install a strong foundation in statistical methods, machine learning, and data visualization for effective decision making.

PEO4. To create awareness of ethical practices, data privacy, and industry standards in data analysis and business intelligence.

PEO5. To enhance professional skills such as communication, teamwork, and leadership, preparing students for roles in data science and analytics within organizations.

PEO6. To equip students with a strong knowledge base for research and innovation in data science, artificial intelligence, and business analytics.

PEO7. To prepare students for successful careers in data science, business analytics, and related fields, enabling them to make significant contributions to the growth of organizations and industries.

Programme Outcomes

PO1. Understanding Core Scientific Concepts: Students will gain clear insight and understanding to recall key scientific principles across various fields. A well established foundational knowledge of the subject will play a crucial role for deeper learning and future studies.

PO2. Commit to Lifelong Scientific Learning: Students will cultivate a habit of continuous learning and shall learn to stay updated with the latest scientific and technological advancements. This mindset will ensure that they remain relevant, engaged and informed throughout their future academic journey.

PO3. Abilities to Analyze and Evaluate: Students will learn to classify and scrutinize complex problems into manageable parts, critically analyse data, and evaluate potential solutions to scientific problems.

PO4. Assessing Ethical Implications: Students will be trained to evaluate the ethical dimensions of research and technological innovations, ensuring that their decisions consider societal impacts and they adhere to ethical standards. This is vital for responsible and sustainable practices.

PO5. Design Experiments and Innovate: Students will learn to design and conduct experiments, developing innovative solutions to challenges through Research Projects. They will also learn to evaluate their results and refine their experimental approaches over time.

PO6. Application of Scientific and Technical Knowledge to Real-World Problems: Students will use their scientific and Technical knowledge and expertise to identify and solve real-world problems. This would involve applying theoretical concepts to practical situations, bridging the gap between classroom learning, Industry-academia and real-life applications.

PO7. Communicating Scientific Findings Effectively: Students will develop the ability to communicate scientific information clearly and effectively, both in writing and verbally. Whether presenting research findings or writing technical reports, clear communication is key to knowledge sharing and collaboration.

PO8. Foster an Interdisciplinary Approach: Students will cultivate leadership and teamwork skills, enabling them to collaborate effectively in diverse, interdisciplinary teams. Leadership qualities such as decision-making and delegation will help them achieve successful outcomes in various projects.

PO9. Promote Environmental Sustainability: Students will understand the environmental impact of scientific activities and advocate for sustainable practices. By considering environmental factors in their work, they will contribute to the long-term health of the planet.

PO10. Enhanced Skills for Entrepreneurship and Employability: Students will be equipped with essential skills for entrepreneurship and employability, focusing on job readiness, soft skills, and practical business knowledge.

Hands-on experience through internships and mentorship will further enhance their readiness for the job market and entrepreneurial ventures.

Programme Specific Outcomes:

PSO1. Explore technical comprehension in statistical analysis and data visualization techniques to effectively interpret and communicate insights from data.

PSO2. Comprehend, explore, and implement machine learning algorithms and data mining techniques for the efficient development of predictive models that address real-world business challenges.

PSO3. Build proficiency in programming languages such as Python, R etc., utilizing them for data manipulation, analysis, and the development of data-driven solutions.

PSO4. Understand and apply database management concepts, including SQL and NoSQL technologies, to ensure efficient data storage and retrieval.

PSO5. Develop the ability to create compelling data visualizations and narratives using tools like Tableau or Power BI, effectively conveying analytical findings to diverse stakeholders.

PSO6. Collaborate within interdisciplinary teams to arrive at solutions to complex problems and develop various analytical and predictive models using project management techniques.

PSO7. Evaluate ethical considerations and data privacy issues related to data collection, storage, and analysis, ensuring responsible use of data in business contexts.

Course Structure

Semester – V			
Course Code	**Course Type**	**Course Name**	**Credits**
	Major	Machine Learning	3
	Major	Data Wrangling	3
	Minor	IOT and Block Chain	2
	Discipline Specific Elective	Advanced Optimization and Simulation Techniques OR Structural Equation Modelling using SPSS AMOS/FOSS software	3
	Vocational	Introduction to Six Sigma	2
	Project		4
	Discipline Specific practical	Machine Learning Practical	1
		Data Wrangling Practical	1
		Advanced Optimization and Simulation Techniques OR Structural Equation Modelling using SPSS AMOS/FOSS software	1
	IKS	IKS for Data Science	2
Total Credits			22
Semester – VI			
Course Code	**Course Type**	**Course Name**	**Credits**
	Major	Categorical and Text Analytics	3
	Major	Artificial Intelligence	3
	DSE	Time Series Analytics OR Marketing Analytics	3
	Vocational	Data Science Case studies	3
	Internship		4
	Discipline Specific Practical	Categorical and Text Analytics Practical	1
		Artificial Intelligence Practical	1

		Time Series Analytics Practical OR Marketing Analytics Practical	1
		Data Science Case studies Practical	1
	IKS	IKS for Data Science	2
Total Credits			22

Detailed Syllabus

Semester V

Major Courses

Curriculum Topics along with Self-Learning topics

Course Name: Machine Learning		**Course Code:**	
Session Per Week(1 session is 60 minutes)		**3**	
Credits		**3**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	**2**	**60**
	Internal	**-**	**15**
	Practical Examination	**2**	**25**

Course outcomes:

1. Understanding the Fundamentals of Machine Learning
2. Implementing Clustering Techniques for Data Segmentation
3. Applying Reinforcement Learning Techniques
4. Building and Evaluating Regression Models
5. Implementing Supervised Learning Algorithms for Categorical Data
6. Understanding and Implementing Advanced Machine Learning Algorithms (CART, C5.0, and CHAID) and apply them for prediction tasks.

Unit	**Content**	**No. of lectures**
1	**Clustering with DBSCAN: From Basics to Advanced Applications:** **1.1 Clustering Basics-** Learn the fundamentals of clustering, focusing on DBSCAN, which groups similar data and detects outliers without needing to specify the number of clusters. **1.2 Tuning DBSCAN-** Discover how to set DBSCAN's key parameters (**eps** and **min_samples**) and use Python to implement and test it on real data. **1.3 Advanced Tips & Applications-** Explore ways to optimize DBSCAN, handle complex data, and apply it in real-world scenarios like customer segmentation and anomaly detection.	**15**
2	Regression Models and Machine Learning Techniques Regression Analysis: Linear Regression. Multiple Regression. Polynomial Regression. Building and interpreting regression models.	**15**

	2.1 Logistic Regression: Introduction to Logistic Regression for classification tasks. Model building and interpretation **2.2** Simple Machine Learning Models: Preprocessing and splitting data. Building basic machine learning models (e.g., KNN, SVM) **2.3** Decision Trees: Understanding decision trees for classification and regression problems. Model building and interpretation **2.4** Model Selection and Cross Validation: Techniques for selecting the best machine learning model. Cross-validation methods for assessing model performance **2.5** Feature Engineering: Techniques for selecting and transforming features to improve model accuracy. Importance of feature scaling and encoding	
3	Advanced Machine Learning Models and Projects Random Forest: Understanding Random Forest as an ensemble learning method. Building and interpreting Random Forest models for classification and regression **3.1** Boosting: Introduction to boosting techniques (e.g., AdaBoost, Gradient Boosting). Building and interpreting boosting models **3.2** Advanced Project Implementation: Application of advanced models such as Random Forest and Boosting. Implementing feature engineering and model selection techniques in real-world projects. Preparing and presenting machine learning solutions for business problems	**15**

Self – Learning Topics (Unit wise)

Sub Unit	Topics
1	Basic Statistics: Mean, Median, Mode, Standard Deviation and Variance.
2	Decision Trees: Understanding decision trees for classification and regression problems
3	Preparing and presenting machine learning solutions for business problems

Online Resources:

"Data Analytics with Python" by PROF. A RAMESH, Department of Management Studies, IIT Roorkee https://nptel.ac.in/courses/106/107/106107220/ for 1.4 and 2.1
"NOC:Data Mining" by Prof. Pabitra Mitra, Computer Science and Engineering, IIT Kharagpur https://nptel.ac.in/courses/106/105/106105174/ for 1.1 and 2.2

Course Name: Machine Learning Practical	**Course Code:**
Sessions Per Week(1 session is 60 minutes)	**2**
Credits	**1**

List of Practical: Machine Learning Practicals	
1	Data Cleaning & Exploration – Load dataset, handle missing values, outliers, and perform basic EDA.
2	Linear & Multiple Regression – Build and evaluate regression models using metrics like R^2 and RMSE.
3	Logistic Regression for Classification – Apply logistic regression, evaluate with accuracy, precision, recall.
4	Decision Tree vs. Random Forest – Compare both models on the same dataset; analyse performance and overfitting.
5	Decision Tree vs. Random Forest – Compare both models on the same dataset; analyse performance and overfitting.
6	Boosting (XGBoost) – Train and tune an XGBoost model; interpret results and compare performance.

Reference books:

1	Max Kuhn and Kjell Johnson, Applied Predictive Modeling
2	Lantz, B (2013), Machine Learning with R, 2nd Ed, PACKT Open Source
3	Miller, J. D. and Forte, R. M. (2015), Mastering Predictive Analytics with R, 2nd Ed, PACKT Open Source.
4	Babcock, J. (2016), Mastering Predictive Analytics with Python, PACKT Open Source.

Course Name: Data Wrangling		**Course Code:**	
Sessions Per Week(1 session is 60 minutes)		**3**	
Credits		**3**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	**2**	**60**
	Internal	**-**	**15**
	Practical Examination	**2**	**25**

Course outcome:

1. List & define various data sources.
2. Critically reflect upon different data sources, types, formats, and structures.

3. Research, justify and apply data cleaning, pre-processing, and standardization for data analytics.
4. Apply data integration concepts and techniques to heterogeneous and distributed data.
5. Interpret, assess, and discuss data quality measurements.
6. Research and justify advanced data wrangling, data integration, and database techniques as relevant to data analytics.

Unit	Content	No. of Lectures
1	Foundations of Data Wrangling & Handling Various Data Formats What is Data Wrangling? Importance & Tasks 1.1 Data Wrangling Tools 1.2 Introduction to Python & Python Basics 1.3 Data meant to be read by machines: CSV, JSON, parQad formats 1.4 Working with External Data Sources: Installing Python packages, Parsing Excel Files, Parsing PDFs (pdfminer) 1.5 Introduction to Data Storage: Databases overview, Relational (MySQL, PostgreSQL, Non-relational (NoSQL), When to use simple file.	**15**
2	Data Cleaning, Preparation, and Exploration: Importance of Data Cleaning, Data Exploration, Validation, and Cleaning: Techniques for exploring and validating data. Handling missing values and outliers. Addressing imbalanced data in datasets. Data cleaning and preprocessing techniques 2.1 Cleaning basics: Formatting, identifying bad data, outliers, duplicates 2.2 Fuzzy Matching & RegEx 2.3 Normalization and Standardization 2.4 Scripting the cleanup & testing with new data 2.5 Data Exploration: Importing, joining datasets, Table operations, Correlation analysis, Grouping and slicing data 2.6 Introduction to Visualization: Charts, time-series, maps, interactive elements	**15**
3	Advanced Visualization & Web Scraping Visual Storytelling: Using words, images, video. Presentation tools & platforms 3.1 Introduction to Web Scraping: What to scrape, network inspection, Web page analysis, interacting with JavaScript 3.2 Web Scraping Tools & Techniques: Reading pages with LXML, XPath, Browser-based scraping: Selenium, Ghost, Spidering with Scrapy, Crawling full websites	**15**

Self-Learning topics (Unit wise)

Sub Unit	Topics
1	Introduction to Python & Python Basics, Databases overview, Relational (MySQL), Installing Python packages.
2	Cleaning basics: Formatting, identifying bad data, outliers, duplicates
3	Visual Storytelling: Using words, images, video. Presentation tools & platforms

Online Resources

Videos on Data Wrangling by Dr. M V Krishna Rao from Institute of Aeronautical Engineering https://www.youtube.com/watch?v=mg7PgYZYBxY&list=PLzkMouYverAK5TE58WLnNq69B4NDRacdb
Data Wrangling with python by Coursera https://www.coursera.org/specializations/data-wrangling-python
Web Automation and Web Scraping by Udemy https://www.udemy.com/course/web-automation-scraping-python/?utm_source=adwords&utm_medium=udemyads&utm_campaign=Search_DSA_Alpha_Prof_la.EN_cc.India&campaigntype=Search&portfolio=India&language=EN&product=Course&test=&audience=DSA&topic=Python&priority=Alpha&utm_content=deal4584&utm_term=_._ag_160270535025_._ad_696202838295_._kw__._de_c_._dm__._pl__._ti_dsa-1705455366924_._li_1007785_._pd__._&matchtype=&gad_source=1&gbraid=0AAAADROdO13bgp6PwUEmOpneHwjjIszi&gclid=Cj0KCQjw_JzABhC2ARIsAPe3ynp8RfGZX2h9NskJZ-8zKNc6joFD4HbXUphVU0M0097Spztqh8i1_IQaAhklEALw_wcB&couponCode=IND21PM

Course Name: Data Wrangling Practicals	**Course Code:**	
Sessions Per Week(1 session is 60 minutes)	**2**	
Credits	**1**	

List of Practical: Data Wrangling Practicals	
1	Import and clean data from CSV, JSON, and XML formats using Python libraries like Pandas.
2	Connect to a MySQL/PostgreSQL database and perform basic queries to import/export data.
3	Write a Python script to clean a dataset by handling outliers, missing values, and normalization.

4	Perform correlation analysis, group/slice data, and visualize results with Python.
5	Create visualizations like bar charts, time-series plots, and maps using Matplotlib or Seaborn.
6	Scrape data from a website using BeautifulSoup, Selenium, or Scrapy and store it for analysis.

Reference books

1	McKinney, Wes. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly Media, 2017.
2	Kazil, Jacqueline & Jarmul, Katharine. Data Wrangling with Python: Working with Unstructured Data in Python. O'Reilly Media, 2016.
3	Beaulieu, Alan. Learning SQL: Master SQL Fundamentals. O'Reilly Media, 2009.
4	Mitchell, Ryan. Web Scraping with Python: Collecting Data from the Modern Web. O'Reilly Media, 2015.

Course Name: Introduction to Internet of Things(IoT) and Block Chain		**Course Code:**	
Sessions Per Week(1 session is 60 minutes)		**2**	
Credits		**2**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	**1**	**30**
	Internal	**-**	**20**

Course outcome:

1. Understand the fundamentals of IoT, IoT architectures, and protocols.
2. Analyze IoT analytics challenges, networking protocols, and cloud analytics.
3. Visualize data and create dashboards using Power BI/Grafana/Kibana/R tools.
4. Apply machine learning, feature engineering, and forecasting methods in IoT data.
5. Evaluate IoT security threats, authentication, and blockchain technologies.
6. Explore blockchain architecture, protocols, and consensus mechanisms.

Unit	Content	No. of lectures
1	**Introduction to Internet of Things(IoT) -I** 1.1 **Introduction to IOT:** Definition of IoT, applications, IoT architectures, introduction to analytics, IoT analytics challenges, IoT devices, Networking basics 1.2 **IOT Protocols:** IoT networking connectivity protocols, IoT networking data messaging protocols, Analyzing data to infer protocol and device characteristics	**15**

	1.3 **IoT Analytics for the Cloud:** Introduction to elastic analytics, Decouple key components, Cloud security and analytics, Designing data processing for analytics, Applying big data technology to storage. **Introduction to Internet of Things(IoT)-II** 1.4 **Visualizing data:** Power BI/ Grafana/ Kibana/ GGSTATSPLOT of R Package for Visualization and Dashboard Making. 1.5 **Data Science for IoT Analytics:** Introduction to Machine Learning, Feature engineering with IoT data, Validation methods, Understanding the bias–variance tradeoff, Use cases for deep learning with IoT data. Time series forecasting using random forest, Time series forecasting using Long Short term Memory Model. 1.6 **Strategies to Organize Data for Analytics:** Linked Analytical Datasets, Managing data lakes, data retention strategy.	
2	**Introduction to Block Chain** **2.1 Introduction to Block Chain:** History, Digital Money to Distributed Ledgers, Design Primitives, Protocols, Security, Consensus, Permissions, Privacy , Block chain Architecture, **2.2 Design and Consensus:** Basic crypto primitives: Hash, Signature, Hash chain to Block chain, Basic consensus mechanisms, Requirements for the consensus protocols, PoW and PoS, Scalability aspects block chain protocols.	**15**

Self – Learning Topics (Unit wise)

Sub Unit	Topics
1	IoT networking connectivity protocols, IoT networking data messaging protocols, Analyzing data to infer protocol and device characteristics GGSTATSPLOT of R Package for Visualization and Dashboard Making.
2	GGSTATSPLOT of R Package for Visualization and Dashboard Making. Introduction to Machine Learning Time series forecasting using random forest
3	History, Digital Money to Distributed Ledgers

Online Resources:

GGSTATSPLOT Presentation by Dr Indrajeet Patil: https://indrajeetpatil.github.io/ggstatsplot/
https://analyticsindiamag.com/ai-trends/how-can-times-series-forecasting-be-done-using-random-forest/
https://www.cse.wustl.edu/~jain/cse570-15/ftp/iot_prot/
https://www.cse.wustl.edu/~jain/cse570-15/ftp/iot_prot/
https://nptel.ac.in/courses/106106139

https://nptel.ac.in/courses/103106123
https://nptel.ac.in/courses/106104220

Reference books:

1.	Minteer, Andrew, Analytics for the Internet of Things (IoT), Packt Publishing Ltd.
2.	Kai Hwang, Min Chen, Big-Data Analytics for Cloud, IoT and Cognitive Computing, Wiley
3.	Hwaiyu Geng, Internet of Things and Data Analytics Handbook, Wiley
4.	John Soldatos, Building Blocks for IoT Analytics Internet-of-Things Analytics, River Publishers
5.	Gerardus Blokdyk, IoT Analytics A Complete Guide, 5starcooks

Discipline Specific Elective (DSE)

Course Name: Advanced Optimization and Simulation Techniques		**Course Code:**	
Session Per Week(1 session is 60 minutes)		**3**	
Credits		**3**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	**2**	**60**
	Internal	**-**	**15**
	Practical Examination	**2**	**25**

Course outcome:

1. Design and implement Monte Carlo simulations for decision-making and risk analysis.
2. Solve non-linear optimization problems using Python/R.
3. Model and simulate systems like queues and resource allocation.
4. Apply gradient-based methods for unconstrained and constrained optimization.
5. Construct PERT/CPM networks and perform project scheduling and critical path analysis.
6. Validate simulation models and ensure the accuracy of generated random variables.

Unit	Content	No. of lectures
1	**1.1 Introduction to Optimization:** Covers the basics of optimization in machine learning, including objective functions, decision variables, and key concepts like local vs global optima and convex vs non-convex optimization.	**15**

	1.2 Gradient Descent & Variants: Focuses on Gradient Descent (GD) and its variants like Stochastic Gradient Descent (SGD), along with advanced techniques like Adam and Nesterov acceleration for efficient optimization. **1.3 Newton-Raphson & Quasi-Newton Methods:** Explores second-order optimization techniques such as Newton-Raphson and quasi-Newton methods (BFGS, L-BFGS), addressing their applications and limitations in large-scale models. **1.4 Stochastic Optimization & QSA:** Covers stochastic optimization methods, including QSA (Quadratic Stochastic Approximation), used for noisy or uncertain optimization problems in machine learning and time-series forecasting.	
2	**Simulation (Python/R)** Introduction to Simulation: Types of simulations: Discrete-event vs. continuous, Random number generation and probability distributions 2.1 Monte Carlo Simulation: Application in decision-making and risk analysis, Implementing in Python/R (e.g., NumPy, simr) 2.2 Discrete Event Simulation (DES): Modeling queues and resource allocation, Using Python (SimPy) and R 2.3 Basic Simulation Techniques: Generating random variables, Validation and verification of models.	**15**
3	**Non-linear Programming (NLP) Problems using Python/R** Introduction to NLP: Convex and concave functions, constraints 3.1 Optimization Techniques: Gradient-based methods (e.g., Steepest Descent, Newton's Method), Unconstrained and constrained optimization 3.2 Solving NLP Problems: Using Python (SciPy) and R (optim, nloptr) 3.3 Applications: Engineering, economics, and operations research	**15**

Self – Learning Topics (Unit wise)

Sub Unit	Topics
1	Discrete-event vs. continuous, Random number generation
2	Convex and concave functions, Applications: Engineering, economics, and operations research
3	Applications in Real-World Projects: Case studies in construction, IT, etc

Online Resources:

Optimization Methods by Dr. D. Nagesh Kumar from IISc Bangalore available on the Swayam portal https://nptel.ac.in/courses/105108127
Introduction to Operations Research by Prof. G. Srinivasan from IIT Madras available on the Swayam portal https://nptel.ac.in/courses/110106062

Course Name: Advanced Optimization and Simulation Techniques Practicals	**Course Code:**
Sessions Per Week(1 session is 60 minutes)	**2**
Credits	**1**

List of Practical: Advanced Optimization and Simulation Techniques Practicals	
1	Model risk and uncertainty in business decisions using random sampling and analyze outcomes.
2	Simulate a queuing system to analyze performance metrics like waiting time and queue length.
3	Implement gradient descent to solve an unconstrained optimization problem and analyze the convergence.
4	Use Python (SciPy) or R (optim) to solve a constrained NLP problem, such as resource allocation in operations research.
5	Construct a PERT/CPM network for a project, identify the critical path, and calculate the project's duration.
6	Create a PERT chart for a project, estimate task durations, and determine the critical path and project completion time.

Reference books:

1	"Introduction to the Theory of Statistics" by Mood, Graybill, and Boes - Great for understanding statistics for simulations.
2	"Convex Optimization" by Stephen Boyd and Lieven Vandenberghe - An excellent reference for NLP.
3	Simulation Modeling and Analysis by Averill M. Law
4	Operations Research: An Introduction by Hamdy A. Taha
5	Project Management: A Systems Approach to Planning, Scheduling, and Controlling by Harold Kerzner

Course Name: Structural Equation Modelling using SPSS AMOS/FOSS software	**Course Code:**
Sessions Per Week(1 session is 60 minutes)	**3**
Credits	**3**

		Hours	Marks
Evaluation System	**Theory Examination**	**2**	**60**
	Internal	**-**	**15**
	Practical Examination	**2**	**25**

Course outcome:

1. Apply principles of SEM for model development, identification, and specification.
2. Perform Exploratory and Confirmatory Factor Analysis (EFA & CFA) to validate factor structures.
3. Integrate mediating and moderating variables in SEM to explore complex relationships.
4. Interpret model fit indices and assess validity through SEM diagnostics.
5. Use statistical software (e.g., AMOS, SPSS, R) for SEM, EFA, and CFA analysis.
6. Implement advanced SEM techniques like mediation, moderation, and mediated moderation.

Unit	Content	No. of lectures
1	Basics of Structural Equation Modelling (SEM): 1.1 Introduction: measurement and structure models, variables and constructs, modelling strategies, conceptualization 1.2 Structural Equation Modelling: Six stages in Model Development, 1) Identification 2) Specification 3) Research design and related issues 1.3 SEM Model Estimation Measurement: Model Structure, Different Estimation Techniques, Issues of Identification 1.4 Model Validity Measurement Mode: Goodness of Fits (GoFs) 1.5 Model validity: Structural models GoFs, Competitive fit, Comparing Models	**15**
2	Exploratory and Confirmatory Factor Analysis: 2.1 conceptualization, Difference between exploratory & confirmatory factor analysis	**15**

	2.2 Objective of EFA, EFA model & assessing measurements, Model validity with Case Study 2.3 Objective of CFA, CFA model & assessing measurements, Model validity, Case Study 2.4 Case study on Multivariate Multiple Regression 2.5 Developing Path Diagrams, Developing Overall Models and identifying Issues, Key Decision Area: Identification and Estimation Interpretation, Model Validity: 4 types of validity, Bootstrapping, Model Diagnostics	
3	3.1 Estimating SEM- CFA Model using MS Excel and Solver: demonstrating solution on MS Excel and solver for conceptual clarity 3.2 Incorporating control variables in SEM- Modelling: Interaction Effects 3.3 Mediation Analysis in SEM: incorporating mediating variables, Case Study 3.4 Moderation Analysis in SEM: incorporating moderating variables, Case Study 3.5 Mediated Moderation Analysis: incorporating both, Case Study	**15**

Self – Learning Topics (Unit wise)

Sub Unit	Topics
1	Introduction: measurement and structure models, variables and constructs, modelling strategies, conceptualization
2	Objective of EFA, EFA model & assessing measurements, Model validity, Case Study
3	Mediation Analysis in SEM: incorporating mediating variables with Case Study

Online Resources:

'Applied Multivariate Statistical Modeling' by PROF. J. Maiti, Department of Mathematics, IIT Kharagpur, available on the NPTEL portal https://nptel.ac.in/courses/111/105/111105091/ for unit II
Structural Equation Modelling (SEM) by Dr. Suresh Sharma Day 1 - https://youtu.be/uWE1rChJtOs Day 2 - https://youtu.be/2wniJL8M1ZQ and Day 3 - https://youtu.be/2VGIKmOZu9g

Course Name: Structural Equation Modelling using SPSS AMOS/FOSS software Practicals	**Course Code**
Sessions Per Week(1 session is 60 minutes)	**2**
Credits	**1**

List of Practical: Structural Equation Modelling using SPSS AMOS/FOSS software Practicals	
1	Develop and specify measurement and structural models, identifying relationships between variables.
2	Perform EFA to extract underlying factors and assess model validity using software like SPSS or R.
3	Conduct CFA to validate hypothesized factor structures and evaluate model fit.
4	Incorporate a mediating variable into SEM to test indirect effects and analyze mediation.
5	Integrate a moderating variable into SEM to assess how it influences the relationship between independent and dependent variables.
6	Perform analysis with both mediating and moderating variables to explore complex relationships in SEM.

Reference books:

1	Mike W.L.Cheung, Meta Analysis: A structural equation modeling Approach,Wiley
2	Rex B. Kline(2011), Principles and Practice of Structural Equation Modeling, Third Edition, TheGuilford Press, New York London
3	Joseph F. Hair Jr. William C. Black Barry J. Babin Rolph E. Anderson, Multivariate Data Analysis, Pearson New International Edition, 7th Edition

Vocational Course

Course Name: Introduction to Six Sigma		**Course Code:**	
Session Per Week(1 session is 60 minutes)		**2**	
Credits		**2**	
		Hours	**Marks**
Evaluation System	**Internal (Practicals)**	**-**	**50**

Course outcome:

1. Define Six Sigma Concepts: define Six Sigma, Lean, and DMAIC, and distinguish between the principles of Lean and Six Sigma.

2. Construct SIPOC and Project Charters: create SIPOC diagrams, define problem and goal statements, and develop project charters.
3. Apply Measurement System Analysis: perform Gage R&R and calculate DPMO, DPU, and process yield.
4. Calculate and Interpret Process Capability: calculate Cp and Cpk indices and assess process capability.
5. Analyze Data Using Statistical Tools: apply regression, correlation, and probability distributions to analyze process data.
6. Use Control Charts for Process Monitoring: construct and interpret X-bar, R, P, and C charts to monitor and control process variation.

Unit	Content	No. of Lectures
1	**Introduction to Six Sigma** 1.1 Quality concept and definition, Six – Sigma: Definition of six-sigma, Principles and Meaning of Lean- six – sigma, Introduction to Lean: Definition of Lean , 5 S in Lean, The 5 Principles of Lean,7 Wastes in Lean . Over view of DMAIC. 1.2 Define Phase: Introduction, Method and its importance, Definitions of VOC,VOP,VOB,CTQ,COPQ with Case study 1.3 Process map, Problem statement , Goal statement, SIPOC, **process components in SIPOC, Mapping the process, Steps to** draw SIPOC,Case study Project charter , Brainstorming, Negative Brainstorming, Case study	**15**
2	**Measure and Analyse** Measure phase: Measurement System Fundamentals, Accuracy , Precision, Repeatability, Reproducibility, Measurement System Analysis: Gage R & R ,Process performance. DPMO, DPU, Yield Process Capability Analysis: Introduction to process capability, concept, Specification limits natural tolerance limits and their comparisons, estimate of percent defectives Capability ratio and Capability indices (Cp), Capability performance indices Cpk with respect to machine and process interpretation, relationship between (i) Cp and Cpk (ii)Defective parts per million and Cp. Analyse Phase : Basic statistics : Types of Data , Descriptive statistics correlation and simple linear regression, Residuals, R^2, adjusted R^2, probability distributions. Process Door and Data Door, Root cause analysis. **Graphs: Histogram, Box – plot, scatter plot , Pareto chart, Run chart, Fishbone diagram (Cause-Effect Diagram).** Improve and Control Improve Phase : Multi Voting , Delphi Technique , Nominal Group Technique , Kaizen, SCAMPER. Introduction, Chance and assignable causes, Statistical basis of the control chart:	**15**

	Basic principles of control chart (Shewhart control charts), Concept of Rational subgroups, Analysis of patterns on control charts, Sensitizing rules for control chart. Control chart for variables: X-Bar, R, S [sample standard deviation] (Construction, charts based on standard values, Interpretation, Applications of variable control charts. In addition problems involving setting up standards for future use is also expected). Introduction to the concept of attribute, Defect. P, np, c and u charts, their uses.(Construction, charts based on standard values, Interpretation, Applications of variable control charts. In addition problems involving setting up standards for future use is also expected). Construction of Charts using R- software/ Python. **Six Sigma Case Study**	

Self-Learning topics (Unit wise)

Sub Unit	Topics
1	Definitions of VOC, VOP, VOB, CTQ, COPQ, process components in SIPOC, Mapping the process, Steps to draw SIPOC, Brainstorming
2	Measurement System Analysis: Gage R & R, Process performance. DPMO, DPU, Yield Analyse Phase: Basic statistics: Types of Data, Descriptive statistics correlation and simple linear regression, Residuals, R^2, adjusted R^2, probability distributions.
3	Six Sigma Case Study

Online Resources

NOC:Six Sigma, IIT Kharagpur by Prof. Jitesh J Thakkar https://nptel.ac.in/courses/110105123 Six Sigma, IIT Kharagpur by Prof. Tapan P. Bagchi https://nptel.ac.in/courses/110105039

Reference books:

1	Fundamental of Mathematical Statistics, Gupta and Kapoor.
2	Probability and Random process by T. Veerarajan.
3	The Certified Six Sigma Green Belt Handbook, Roderick A. Munro and Govindarajan Ramu , American Society for Quality (ASQ)
4	What Is Design For Six Sigma,(2005), Roland Cavanagh, Robert Neuman, Peter Pande, Tata McGraw- Hill

5	Introduction to Statistical Quality Control(2009), Montgomery, Douglas, C ,Sixth Edition, John Wiley & Sons.Inc.:.
6	Statistical Quality Control: E.L.Grant. 2nd edition, McGraw Hill, 1988.
7	Quality Control and Industrial Statistics: Duncan. 3rd edition, D.Taraporewala sons & company.
8	Quality Control: Theory and Applications: Bertrand L. Hansen, (1973),Prentice Hall of IndiaPvt. Ltd..
9	Introduction to Statistical Quality Control(2009), Montgomery, Douglas, C. , Sixth Edition, John Wiley & Sons, Inc.:
10	Quality Control (1976), I.V. Burr, Mardekkar, New York,
11	Fundamentals of Applied Statistics , Gupta and Kapoor

IKS

Course Name: IKS for Data Science		**Course Code:**	
Sessions Per Week(1 session is 60 minutes)		**2**	
Credits		**2**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	**1**	**30**
	Internal	**-**	**20**

Course outcome:

1. Describe the core components and philosophy of Indian Knowledge Systems and their interdisciplinary nature. (Understand)

2. Identify connections between traditional Indian knowledge domains and modern scientific thinking. (Apply)

3. Critically evaluate the epistemological methods of IKS and their relevance to contemporary research. (Analyze)

4. Explain ancient Indian contributions in mathematics, astronomy, logic, and linguistics. (Understand)

5. Compare Indian logical systems (Nyaya, Paninian grammar) with modern computing and AI principles. (Analyze)

6. Demonstrate how ancient Indian algorithms and classification methods can inform modern data models. (Apply)

Unit	Content	No. of lectures
1	**Introduction to Indian Knowledge Systems (IKS)** **Overview of IKS**: Definition, scope, and relevance in modern times.	**15**

	Major Disciplines: Ayurveda, Mathematics, Astronomy, Logic, Linguistics, Agriculture, Architecture, and Music. **Philosophical Foundations**: Nyaya, Vaisheshika, and Sankhya systems. **IKS and Scientific Temper**: Epistemology and classification of knowledge in Indian traditions. **Connection to Data Science**: Importance of structured knowledge, observation, inference, and categorization.	
2	Ancient Indian Contributions Relevant to Data Science **Indian Mathematics**: Contributions of Aryabhata, Bhaskara, and Brahmagupta. Concepts of zero, infinity, algorithms, and combinatorics (e.g., Pingala's Chandaḥśāstra). **Linguistics & Panini's Grammar**: Panini's Ashtadhyayi as an early formal system (precursor to formal grammars and rule-based AI). **Logic Systems**: Nyaya Sutras: Indian logic, reasoning, and inference (Anumana) – parallels with modern logic and AI. **Data Classification in Ancient Texts**: Taxonomy in Ayurveda and ancient Indian botany.	**15**

Online Resources:

Bharatiya Jnana Parampara (Indian Knowledge System) – AICTE Module, AICTE IKS Division, https://iksindia.org
Indian Knowledge System MOOC, Prof. Ganti S. Murthy, SWAYAM, https://swayam.gov.in
National Mission on Education through ICT – IKS, MHRD, NMEICT Portal, http://www.iksindia.org
Mathematics in India: From Vedic Period to Modern Times, TIFR Lecture Series, YouTube/TIFR, YouTube Link
Panini and Computational Linguistics, Prof. Rajeev Sangal, IIIT-H Talks, https://sangal.in/panini
Nyaya and Indian Logic: Stanford Encyclopedia, Jonardon Ganeri, https://plato.stanford.edu/entries/logic-india/
Digital Library of India – Ancient Texts, MHRD/IIIT-H, https://www.dli.gov.in

Reference books:

1	Indian Knowledge Systems, Kapil Kapoor, Indian Institute of Advanced Study (IIAS)
2	Indian Knowledge System: Concepts and Applications, B.L. Atrey & S.R. Bhatt, Prabhat Prakashan
3	An Introduction to Indian Philosophy, Satischandra Chatterjee & Dhirendramohan Datta, University of Calcutta Press
4	A History of Indian Literature Vol. I: Vedic and Epic Periods, Maurice Winternitz, Motilal Banarsidass
5	The Āryabhaṭīya of Āryabhaṭa, Walter Eugene Clark (Translator), University of Chicago Press
6	Indian Logic: A Reader, Jonardon Ganeri, Curzon Press
7	Panini: His Work and Its Traditions, George Cardona, Motilal Banarsidass

Detailed Syllabus

Semester VI

Major Courses

Curriculum Topics along with Self-Learning topics

Course Name: Categorical and text analytics		**Course Code:**	
Sessions Per Week(1 session is 60 minutes)		**3**	
Credits		**3**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	**2**	**60**
	Internal	**-**	**15**
	Practical Examination	**2**	**25**

Course outcome:

1. Explain and summarize categorical data, probability distributions, and statistical inference for discrete data.
2. Evaluate contingency tables, calculate risk and odds ratios, and perform tests of independence.
3. Analyze nominal and ordinal associations using Gamma and Eta coefficients and assess relationships between categorical variables.
4. Fit and interpret GLMs, including loglinear and logistic models, for binary and count data, with statistical inference and model checking.
5. Import, clean, and transform text data while implementing visualization techniques like word clouds and comparison clouds.
6. Perform sentiment analysis on individual files and corpora to extract insights and support decision-making.

Unit	**Content**	**No. of lectures**
1	Categorical data: 1.1 Describing categorical data 1.2 Categorical response data, Probability distributions for categorical data. Statistical inference for discrete data. 1.3 Contingency tables: Probability structure for contingency tables, Comparing proportions with 2x2 tables, The Risk ratio and odds ratio, Tests for independence, Exact inference, Extension to three-	**15**

	way and larger tables. Measures of nominal and ordinal Association Gamma and Eta Coefficients), Association between categorical variables	
2	Generalized linear models: 2.1 Loglinear models for contingency tables: 2.2 Loglinear models for two-way and three-way tables, Inference for Loglinear models, The loglinear-logistic connection, Multinomial and Ordinal logistic, Independence graphs and collapsibility. 2.3 Fitting of Logit Models and Loglinear Models. 2.4 Components of a generalized linear model, 2.5 GLM for binary and count data, Statistical inference and model checking, 2.6 Fitting GLMs	**15**
3	Text and Sentiment Analytics: 3.1 Importing Text from a Single File, Multiple Files, Folders, and the Web (Fetching data from Twitter with and /or without an authenticated account.) 3.2 Text Collection and Transformation, Text Cleaning Sequence and Operations, 3.3 Finding Frequent Terms, Word Associations, Frequency Table 3.4 Visualization - Word Cloud and Word Cloud 2 3.5 Comparison Cloud and Common Cloud 3.6 Sentiment Analysis on Individual File 3.7 Sentiment Analysis on Corpus 3.8 Text mining and Modeling. 3.9 Story Telling on a. Extract the data from Twitter regarding omicron variant in India and perform cluster analysis. b. Extract data for your favourite celebrity and perform sentiment analysis c. Extract data from Twitter of top 4 metro cities location and compare the polarity and subjectivity analysis d. Extract the data from IMDB or any other source and build a regression model for predicting the sales for new upcoming movie like 83	**15**

	e. Extract the data from IMDB or any other source and build a classification model to predict if the movie will be hit or flop f. Extract data from your favorite motivational book and performs sentiment analysis and create world cloud	

Self – Learning Topics (Unit wise)

Sub Unit	Topics
1	Describing categorical data, Association between categorical variables
2	Components of a generalized linear model,
3	Text Analytics: Text Collection and Transformation, Text mining and Modeling.X

Online Resources:

"Introduction to probability and Statistics" by Prof. G. Srinivasan, Department of Mathematics, IIT Madras https://nptel.ac.in/courses/111106112
Business Analytics & Text Mining Modelling Using Python, available on the Swayam portal, GOI by Prof. Gaurav Dixit, IIT Roorkee given on week 8/unit 8 https://nptel.ac.in/courses/110/107/110107129/ for unit 3

Course Name: Categorical and text analytics Practicals	**Course Code:**
Sessions Per Week(1 session is 60 minutes)	**2**
Credits	**1**

List of Practical: Categorical and Text Analytics Practicals	
1	Analyze categorical data, create contingency tables, perform Chi-square tests, and calculate odds and risk ratios.
2	Fit a logistic regression model on binary data, assess model performance, and interpret coefficients (odds ratios).
3	Extract tweets about a specific topic (e.g., Omicron), preprocess, and analyze sentiment using VADER/TextBlob.
4	Collect tweets, preprocess, vectorize text, apply clustering (e.g., K-means), and visualize clusters.

5	Create word clouds from text data (e.g., tweets or books) to visualize frequent terms and create comparison clouds.
6	Build a regression model to predict movie sales using features like budget, genre, and cast, and evaluate the model's performance.

Reference books:

1	Cox, D. R. and Snell, E. J. (1989), The Analysis of Binary Data. CRC Press. 2nd Edition.
2	Hosmer, D. W. and Lemeshow, S. (2000), Applied Logistic Regression. John Wiley, 2nd Edition
3	Lior Rokach & Oded Z. Maimon, Data Mining with Decision Trees - Theory & Applications.
4	Agresti, A. (2013), Categorical Data Analysis, 3rd Edition. New York: Wiley (an advanced text).
5	Gokhale, D. V. and Kullback, S. (1978), The Information in Contingency Tables. Marcel Dekker.

Course Name: Artificial Intelligence		**Course Code:**	
Sessions Per Week(1 session is 60 minutes)		**3**	
Credits		**3**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	**2**	**60**
	Internal	**-**	**15**
	Practical Examination	**2**	**25**

Course outcome:

1. Define AI concepts, components, and applications in real-time analytics.
2. Outline the process of implementing AI solutions and understand deep learning basics.
3. Apply automated problem-solving techniques like state-space and heuristic search.
4. Explain neural networks, perceptrons, and deep learning fundamentals.
5. Evaluate deep learning techniques, activation functions, and optimization methods.
6. Analyze AI applications in finance, marketing, and HR using analytics for optimization.

Unit	**Content**	**No. of lectures**

1	**Introduction to Artificial Intelligence and Search Algorithm** What is artificial intelligence, A brief review of AI history , Related research fields, Problem formulation Review of tree structure, Review of graph structure, Graph implementation, State space representation, Search graph and search tree **Simple Search Algorithm - I** - Random search, Search with closed list, Search with open list, Depth-first and breadth-first search again, Uniform-cost search **Simple Search Algorithm-II**- What are heuristics? What is heuristic search? Best first search A* algorithm, Generalization of search problems **Adversarial Search** Minimax Algorithm for two player games , An Example of Minimax Search, An Example of Minimax Search , Analysis of Alpha Beta Pruning , Horizon Effect, Game Databases & Other Ideas	**15**
2	**Reasoning:** Proposition and first-order logic, Rule-based systems, semantic net, conceptual graph, inference and deduction, Resolution refutation, answer extraction, Reasoning under uncertainty - probabilistic reasoning, belief networks **Production system:** Inference engine, Working memory, Knowledge base , Pattern matching , Conflict resolution , Forward inference, Back inference	**15**
3	**Genetic Algorithms:** Difference between traditional algorithms and Genetic Algorithm (GA); Basic concepts of GA; Working principle; Encoding methods; Fitness function; GA Operators: Reproduction, Crossover, Mutation; Convergence of GA; Detailed algorithmic steps; Adjustment of parameters; Multicriteria optimization; Solution of typical problems using genetic algorithm; Recent applications **Hybrid Systems:** Introduction to hybrid systems: Fuzzy-neural systems, Genetic Fuzzy systems, Neuro-genetic systems.	**15**

Self – Learning Topics (Unit wise)

Unit	Topics

1	A brief review of AI history, Related research fields, Review of tree structure, Review of graph structure, Graph implementation, Horizon Effect, Game Databases & Other Ideas
2	Proposition and first-order logic, Rule-based systems, semantic net, conceptual graph
3	Difference between traditional algorithms and Genetic Algorithm (GA), Recent applications

Online Resources:

"An Introduction To Artificial Intelligence" by Prof. Mausam, Department OfComputer Science and Engineering,IIT Delhi https://nptel.ac.in/courses/106/102/106102220/
"Artificial Intelligence : Search Methods For Problem Solving" by PROF. DEEPAK KHEMANI, Department of Computer Science and Engineering, IIT Madras https://nptel.ac.in/courses/106/106/106106126/

Reference books:

1	Artificial Intelligence: A Modern Approach Book by Peter Norvig and Stuart J. Russell
2	S. Rajasekaran and G. A. Vijayalakshmi Pai, Neural Networks, Fuzzy Logic and Genetic Algorithm: Synthesis and Applications, PHI.
3	"Artificial Intelligence: Foundations of Computational Agents" by David L. Poole and Alan K. Mackworth
4	S. N. Sivanandam and S. N. Deepa, Principles of Soft Computing, 2nd ed., Wiley India.

Course Name: Artificial Intelligence Practicals	**Course Code:**
Sessions Per Week(1 period is 60 minutes)	**2**
Credits	**1**

List of Practical: Artificial Intelligence Practicals	
1	Implement Depth-First and Breadth-First Search on a tree/graph structure.
2	Apply A* algorithm for pathfinding in a maze or grid using heuristics.
3	Simulate Minimax algorithm with Alpha-Beta pruning for a two-player game.
4	Build a simple rule-based expert system using first-order logic and inference.
5	Implement Genetic Algorithm for solving optimization problems like the traveling salesman.
6	Develop a hybrid fuzzy-neural system for classification tasks.

Vocational Course

Course Name: Data Science Case Studies		**Course Code:**	
Sessions Per Week(1 session is 60 minutes)		**3**	
Credits		**3**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	**2**	**60**
	Internal	**-**	**15**
	Practical Examination	**2**	**25**

Course outcome:

1. Apply machine learning models like Logistic Regression, SVM, and Random Forest using scikit-learn.
2. Build models for detecting emotions in text and associating emojis.
3. Master NLP techniques such as tokenization, stemming, and lemmatization with NLTK.
4. Develop a hotel recommendation system using customer reviews and ratings.
5. Conduct customer personality analysis for personalized marketing.
6. Utilize Python and R cheat sheets for efficient data science workflows.

Unit	Content	No. of Lectures
1	Data Science Case Studies -I 1.1 **scikit-learn Library in Python** 1.1.1 Logistics Classifier Model 1.1.2 Discriminant Classifier Model 1.1.3 Support Vector Classifier Model : 1.1.4 Random Forest Classifier Model 1.1.5 Decision Tree Classifier Model 1.2 **Case Study on Digital Story Telling and Social Media**	**15**

	1.3 **Case Study on Text Emotions Detection** Humans can express their emotions in any form, such as the face, gestures, speech and text. The detection of text emotions is a content-based classification problem. Detecting a person's emotions is a difficult task, but detecting the emotions using text written by a person is even more difficult as a human can express his emotions in any form. Recognizing this type of emotion from a text written by a person plays an important role in applications such as chatbots, customer support forum, customer reviews etc. Train a machine learning model that can identify the emotion of a text by presenting the most relevant emoji according to the input text.	
2	Data Science Case Studies -II 2.1 **Natural Language Toolkit in Python** 2.1.1 Tokenizing 2.1.2 Filtering Stop Words 2.1.3 Stemming 2.1.4 Tagging Parts of Speech 2.1.5 Lemmatizing Hotel Recommendation System 2.2 **Case Study on** Hotel Recommendation System: The first thing to do when planning a trip is finding a hotel. There are many websites recommending the best hotel for our trip. A hotel recommendation system aims to predict which hotel a user is most likely to choose from among all hotels. Build a type of system which will help the user to book the best hotel out of all the other hotels. We can do this using customer reviews and ratings.	**15**
3	Data Science Case Studies -III 3.1 Cheat Sheet on gto76, Persolim.fr, Pythoncheatsheet.org, WebsiteSetup, Cheatography, Python for Data Science, Real Python, **matplotlib in Python** and **seaborn Library in Python** 3.2 Cheat Sheet of R, Tidyverse, data.table R Package Cheat Sheet, Data Visualization with ggplot2 cheatsheet, RStudio Cheatsheets. **3.3 Case Study on Customer Personality Analysis** Customer Personality Analysis is a detailed analysis of a company's ideal customers. It helps a business to better understand its customers and makes it easier for them to modify products according to the specific needs, behaviours and concerns of different types of customers. Perform an analysis that should help a business to modify its product based on its target customers from different types of customer segments.	**15**

Self-Learning topics (Unit wise)

Sub Unit	Topics
1	**scikit-learn Library in Python:** Support Vector Classifier Model, Random Forest Classifier Model, Decision Tree Classifier Model

2	**Natural Language Toolkit in Python**: Filtering Stop Words, Tagging Parts of Speech, Lemmatizing Hotel Recommendation System
3	**matplotlib in Python** and **seaborn Library in Python**

Online Resources

Online resources available on https://scikit-learn.org/stable/
Online resources available on https://towardsdatascience.com/
Certification course on natural language processing from NPTEL Swayam https://onlinecourses.nptel.ac.in/noc19_cs56/preview
Data science applications for Real-world analysis https://www.youtube.com/live/GJIqGA2G650?feature=shared

Course Name: Data Science Case Studies Practicals	**Course Code:**
Sessions Per Week(1 session is 60 minutes)	**2**
Credits	**1**

List of Practical: Data Science Case Studies Practicals	
1	Implement and evaluate different classifiers (Logistic Regression, SVC, etc.) using scikit-learn.
2	Build an emotion detection model from text and map emotions to emojis.
3	Preprocess text data using NLTK (tokenization, stop word removal, stemming, etc.).
4	Create a hotel recommendation system using customer reviews and ratings.
5	Perform customer personality analysis with K-Means clustering.
6	Visualize data using Matplotlib and Seaborn for insights.

Reference books

1	Data science from scratch by Steven Cooper, Data Science
2	Data science and complex networks : real cases studies with Python: Caldarelli, Guido, Chessa, Alessandro, Oxford University Press
3	Text Emotions Detection with Machine Learning https://thecleverprogrammer.com/2021/02/19/text-emotions-detection-with-machine-learning/
4	Customer Personality Analysis with Python

5	https://thecleverprogrammer.com/2021/02/08/customer-personality-analysis-with-python/
6	Hotel Recommendation System with Machine Learning https://thecleverprogrammer.com/2021/02/13/hotel-recommendation-system-with-machine-learning/
7	Python for Data Science is a one-page Python cheat sheet to learn the fundamentals of this programming language. https://analyticsindiamag.com/top-10-python-cheat-sheets-in-2022/
8	RStudio Cheatsheets https://www.rstudio.com › resources › cheatsheets https://www.google.com/search?client=firefox-b-d&q=r+cheat+sheet

Discipline Specific Elective (DSE)

Course Name: Time Series Analytics		**Course Code:**	
Sessions Per Week(1 session is 60 minutes)		**3**	
Credits		**3**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	**2**	**60**
	Internal	**-**	**15**
	Practical Examination	**2**	**25**

Course outcome:

1. Recall basic concepts in time series analysis and forecasting
2. Compare with multivariate times series and other methods of applications
3. Display knowledge of, and a critical understanding of, the main properties of MA, AR, ARMA, ARIMA, and RW models
4. Distinguish the ARIMA modelling of stationary and nonstationary time series
5. Criticize and judge time series regression models.
6. Simulating time series models and regression models for time series

Unit	**Content**	**No. of Hours**

1	Introduction 1.1 Product Design, Introduction To Forecasting, Definition of time series, Its components. Models of time series. 1.2 Estimation of trend by: i) Freehand curve method ii) method of semi average iii)Method of Moving average iv) Method of least squares(linear trend only) 1.3 Accuracy measurements: Mean absolute percentage error, Root mean square error. 1.4 Estimation of seasonal component by i) method of simple average ii) Ratio to moving average iii) Ratio to trend method.	15
2	Time Series Formulation : 2.1 Story telling on Exponential smoothing: Single, double, Triple using R/ Python Software 2.2 Mathematical Formulation of time series. Time series as a discrete parameter stochastic process. Concept of strict stationary, stationary upto order 'm', Mean stationary, covariance stationary. 2.3 Auto covariance function (ACVF) and its properties. Auto correlation function (ACF) and its properties. Partial auto correlation function (PACF). 2.4 Portmanteau tests for noise sequences, transformation to obtain Gaussian series. Story telling on Exponential smoothing etc by extracting data from yahoo finance for Nifty 50. By typing stock's ticker name in search and click on the "Historical Data" tab. variohttps://finance.yahoo.com/quote/%5ENSEI/history?p=%5ENSEI	15
3	Time Series Models with python from various domains like sales, stock price, Gold price, bitcoins, Kaggle site data sets Story Telling on 3.1 How to Work with Time Series Data with Pandas 3.2 Use Statsmodels to Analyze Time Series Data 3.3 General Forecasting Models - ARIMA(Autoregressive Integrated Moving Average), Auto regressive integrated moving average (ARIMA) models 3.4 General Forecasting Models-SARIMA(Seasonal Autoregressive Integrated Moving Average) General Forecasting Models - SARIMAX	15

Self-Learning topics (Unit wise)

Unit	Topics
1.1	Product Design, Introduction To Forecasting, Definition of time series, Its components.
2.1	Story telling on Exponential smoothing
2.2	Auto covariance function (ACVF) and its properties. Auto correlation function (ACF) and its properties. Partial auto correlation function (PACF).
3.1	Auto regressive (AR), Moving average (MA) and Autoregressive moving average (ARMA)
3.2	Auto regressive integrated moving average (ARIMA) models

Online Resources

"Applied Time Series Analysis" by Prof Arun K. Tangirala, Department of Chemical Engineering, IIT Madras Source: https://nptel.ac.in/courses/103/106/103106123/
"Production and Operation Management" by Prof. Rajat Agarwal, IIT Roorkee available on NPTEL Link: https://nptel.ac.in/courses/110107141 for 1.1 and 2.1

Course Name: Time Series Analytics Practicals	**Course Code:**
Session Per Week(1 period is 60 minutes)	**2**
Credits	**1**

List of Practical: Time Series Analytics Practicals	
1	Problem solving based on Unit 1
2	Estimation of trends using freehand curve and moving average method.
3	Implementing ARIMA model to predict stock prices.
4	Use SARIMA for modeling seasonal data.

5	Forecast using Exponential Smoothing and evaluate using MAPE/RMSE.
6	Apply SARIMAX model for forecasting with external variables

Reference books

1	Forecasting and Control, George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung, Wiley, 2015
2	Introduction to Time Series and Forecasting, Peter J. Brockwell, Richard A. Davis, Springer, 2016
3	Practical Time Series Analysis: Master Time Series Data Processing, Visualization, and Modeling using Python, Aileen Nielsen, O'Reilly Media, 2020
4	Forecasting: Principles and Practice, Rob J. Hyndman, George Athanasopoulos, OTexts, 2021
5	Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, Wes McKinney, O'Reilly Media, 2018

Course Name: Marketing Analytics		**Course Code:**	
Sessions Per Week(1 session is 60 minutes)		**3**	
Credits		**3**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	**2**	**60**
	Internal	**-**	**15**
	Practical Examination	**2**	**25**

Course outcome:

1. Manipulate, clean, and visualize data using R, applying Conjoint Analysis for customer insights.
2. Segment customers using clustering algorithms and conduct Market Basket Analysis for purchasing patterns.
3. Build time series forecasting models and optimize pricing strategies based on demand.
4. Develop and evaluate Marketing Mix Models and create personalized recommender systems
5. Predict customer churn, calculate CLV, and design retention strategies.
6. Perform sentiment analysis on text data and apply social **network analysis to identify influencers.**

Unit	**Content**	**No. of Hours**

1	**Introduction to R Programming and Customer Analysis** R environment setup (installation and packages), Data types and structures in R (vectors, data frames, lists, matrices), Data manipulation (dplyr, tidyr), Basic data visualization using ggplot2, Writing Customer Value and Conjoint Analysis: Defining customer value and customer satisfaction, Introduction to Conjoint Analysis and its application in marketing research, Steps involved in performing Conjoint Analysis, Using R for Conjoint Analysis (choice-based conjoint, utility estimation) Customer Segmentation and Targeting: Overview of customer segmentation and targeting, K-means clustering and hierarchical clustering, Market basket analysis for segmentation,Demographic, psychographic, and behavioral segmentation	**15**
2	**Pricing, Forecasting, and Marketing Models** Demand Forecasting and Pricing: Time series forecasting techniques (ARIMA, Exponential Smoothing), Factors influencing demand (price elasticity, seasonality, promotions), Price optimization models, Introduction to regression models for pricing Advanced Pricing Strategies: Dynamic pricing models, Price discrimination and personalized pricing, Price sensitivity analysis, Implementing pricing strategies using R Marketing Mix Models and Advertising Models: Introduction to Marketing Mix Models (4Ps: Product, Price, Place, Promotion), Regression-based Marketing Mix Models, Advertising effectiveness models, Attribution modeling in marketing Recommender Systems: Types of recommender systems (collaborative filtering, content-based filtering), Building a simple collaborative filtering model in R, Evaluation metrics for recommender systems (precision, recall, F1-score)	**15**
3	**Advanced Marketing Analytics and Customer Retention and Network Analysis** Market Basket Analysis and RFM Analysis: Introduction to Market Basket Analysis (association rules, Apriori algorithm), RFM (Recency, Frequency, Monetary) Analysis for customer behavior segmentation, Using R to perform Market Basket and RFM Analysis Customer Churn and Customer Lifetime Value (CLV): Techniques for churn prediction (logistic regression, decision trees), CLV calculation methods and models, Implementing retention strategies based on churn and CLV analysis. Text Mining and Sentiment Analysis: Introduction to Text Mining, Preprocessing text data (tokenization, stemming, stop words removal), Sentiment Analysis using R, Text classification techniques Text Mining and Product Innovation Management: Identifying trends through text mining, Text mining for product feedback and feature extraction,Linking customer feedback to product innovation strategies	**15**

	Social Network Analysis for Marketing: Overview of social network theory, Centrality measures and influence in networks, Using R for social network analysis (igraph package), Applications of social network analysis in marketing (influencer identification, virality)	

Self-Learning topics (Unit wise)

Unit	Topics
1	R environment setup (installation and packages), Data types and structures in R (vectors, data frames, lists, matrices), Overview of customer segmentation and targeting
2	Introduction to regression models for pricing, Introduction to Marketing Mix Models (4Ps: Product, Price, Place, Promotion)
3	Introduction to Text Mining, Preprocessing text data (tokenization, stemming, stop words removal)

Online Resources

Marketing Analytics Course from SWAYAM portal https://onlinecourses.nptel.ac.in/noc25_mg45/preview

Course Name: Marketing Analytics Practicals	**Course Code:**
Sessions Per Week(1 period is 60 minutes)	**2**
Credits	**1**

List of Practical: Marketing Analytics Practicals	
1	Use R for data manipulation, visualization, and perform Conjoint Analysis on product preference data.
2	Implement customer segmentation using clustering algorithms and conduct Market Basket Analysis in R.
3	Build demand forecasting models using time series analysis and optimize pricing using R.
4	Develop a Marketing Mix Model and build a recommender system using collaborative filtering in R.
5	Implement churn prediction models and calculate CLV in R for customer retention.
6	Perform sentiment analysis on customer reviews and apply social network analysis in R for influencer identification.

Reference books

1	R for Data Science, Hadley Wickham and Garrett Grolemund, O'Reilly Media
2	Marketing Research: An Applied Orientation, Naresh K. Malhotra, Pearson Education

3	Applied Predictive Modeling, Max Kuhn and Kjell Johnson, Springer
4	Marketing Analytics: Strategic Models and Metrics, Stephan Sorger, CreateSpace Independent Publishing
5	Data Mining for Business Analytics: Concepts, Techniques, and Applications in R, Galit Shmueli, Peter C. Bruce, Inbal Yahav, Nitin R. Patel, Wiley

IKS

Course Name: IKS for Data Science		**Course Code:**	
Sessions Per Week(1 session is 60 minutes)		**2**	
Credits		**2**	
		Hours	**Marks**
Evaluation System	**Theory Examination**	**1**	**30**
	Internal	**-**	**20**

Course outcome:

1. Analyze the alignment between ancient Indian taxonomies and modern data ontologies. (Analyze)
2. Apply structured knowledge from Sanskrit and Ayurveda to solve basic problems in NLP or knowledge representation. (Apply)
3. Evaluate ethical frameworks from Indian philosophy and their application in data governance and AI. (Evaluate)
4. Analyze real-world case studies that combine Indian traditional knowledge with data science techniques. (Analyze)
5. Design a mini-project or model integrating traditional datasets (e.g., Ayurveda, agriculture) with modern tools. (Create)
6. Present informed arguments on the feasibility and innovation potential of IKS-driven data solutions. (Evaluate/Create)

Unit	**Content**	**No. of lectures**
1	Integrating IKS with Modern Data Science Tools **Ontology & Knowledge Representation**: How ancient Indian classification systems align with data structures and ontologies. **Natural Language Processing (NLP)**: Sanskrit as a highly structured language; relevance in modern NLP and AI. **Big Data in Ancient Context**: Handling large datasets (e.g., Vedic texts, astronomical calculations). **Ethics in Data Science**:	**15**

	Ethical considerations and sustainable practices from Indian philosophy (e.g., Dharma, Ahimsa).	
2	Case Studies and Applications **Ayurveda and AI**: Data modeling of diagnostic methods and treatment plans. **Vedic Mathematics**: Algorithmic thinking and mental math techniques. **Smart Agriculture**: Traditional agricultural knowledge fused with IoT and data analytics. **IKS and Machine Learning**: Feature selection and categorization inspired by ancient classification systems. **Projects/Presentations**: Group work exploring a dataset with IKS insights (e.g., medicinal plants, classical music patterns, Panchang data analysis).	**15**

Online Resources:

Sanskrit and Natural Language Processing, Amba Kulkarni, JNU/Hyderabad University, http://sanskrit.uohyd.ac.in
IKS and AI: Talk by Dr. Vijay Bhatkar, Vigyan Bharati/YouTube, YouTube Link
NLP Resources for Sanskrit and Indian Languages, TDIL Programme, MeitY Govt. of India, https://tdil.meity.gov.in
Vedic Mathematics Online Course, Bharatiya Vidya Bhavan https://vedicmathsindia.org
Ayurveda Informatics, CCRAS (Central Council for Research in Ayurvedic Sciences), https://ccras.nic.in
Indian Traditional Knowledge Digital Library (TKDL), CSIR-Govt. of India, https://www.tkdl.res.in
Data-Driven Models in Indian Agriculture, ICAR Reports, icar.org.in, https://icar.org.in

Reference books:

1	Formal Grammar and Panini, Peter Scharf, Harvard Oriental Series
2	Natural Language Processing for Sanskrit, Amba Kulkarni (ed.), University of Hyderabad Publications
3	Ontology and the Lexicon: A Natural Language Processing Perspective, Chu-Ren Huang & Nicoletta Calzolari, Cambridge University Press (for modern parallels)
4	Vedic Mathematics, Bharati Krishna Tirtha, Motilal Banarsidass
5	Ayurveda and the Mind, David Frawley, Motilal Banarsidass
6	Ancient Indian Agriculture, Om Prakash, Concept Publishing Company
7	Indian Classical Music and the Gramophone 1900–1930, Vikram Sampath, Roli Books