

**R. P. Gogate College of Arts & Science and
R. V. Jogalekar College of Commerce (Autonomous),
Ratnagiri**



**Master of Science (M.Sc.)(Data Science)
Two Year Integrated Programme-
Four Semesters**

Course: Data Science

Course Structure

Syllabus for MSc Semester III & IV

**Under Choice Based Credit System (CBCS)
To be implemented from Academic Year 2026-2027**

Progressively

Preamble

The Master of Science (M.Sc.) in Data Science is a graduate-level program offered by the autonomous college, focusing on the study and application of data analysis, data mining, machine learning, and statistical modeling techniques to extract valuable insights and knowledge from large and complex datasets. This program is meticulously designed to equip students with the necessary skills and knowledge to tackle the challenges and opportunities in the rapidly evolving field of data science.

The curriculum of the M.Sc. in Data Science program typically includes a blend of core courses and elective courses. Core courses provide a solid foundation in fundamental concepts and techniques of data science, such as programming languages (e.g., Python, R), essential technologies for data science, data visualization, statistical analysis, natural language processing, soft computing, time series analysis and forecasting, and deep neural networks. Elective courses allow students to specialize in specific areas of interest, such as healthcare analytics, sports analytics, legal analytics, human resource and retail marketing analytics, blockchain technologies, image and video analytics, and social media analytics.

Additionally, the M.Sc. in Data Science program incorporates On Job Training (OJT) and hands-on projects. This intensive OJT training, equivalent to a full course, provides invaluable exposure to real-world scenarios within IT or IT-related organizations. By applying their theoretical knowledge in practical settings, students gain firsthand experience and develop the necessary skills to thrive in the professional world. The projects help students gain practical experience in applying their knowledge to solve complex data challenges and provide valuable insights to organizations. The inclusion of a Research Methodology Course helps students develop a strong research attitude, enabling them to contribute meaningfully to the advancement of Data Science.

Career prospects for graduates of the M.Sc. in Data Science program are highly promising, as data science skills are in high demand across various industries. Graduates can pursue careers as data scientists, data analysts, machine learning engineers, data engineers, or data consultants in sectors such as finance, healthcare, e-commerce, marketing, and technology. They may also find opportunities in research institutions or pursue further studies at the doctoral level.

Overall, the M.Sc. in Data Science program offers a comprehensive and rigorous education in the field of data science, equipping students with the skills and knowledge necessary to succeed in a data-driven world. The curriculum's continuous refinement has been made possible through the valuable inputs, suggestions, and observations of colleagues, experts from premier institutions, and industry professionals. The autonomous college extends its gratitude to all those who have directly or indirectly contributed to the development of this program.

Name of Programme	Masters of Data Science
Level	PG
No of Semesters	04
Year of Implementation	2026-27
Programme Specific Outcomes (PSO)	<p>At the end of the Programme, Learner will be able to</p> <ol style="list-style-type: none"> 1. Develop a solid foundation in fundamental concepts, theories, and methodologies of Data Science. 2. Offer opportunities for specialization in a chosen area of Data Science. 3. Foster a research-oriented mindset and contribute to the advancement of Data Science. 4. Prepare learners for lifelong learning, adapting to emerging technologies and industry requirements. 5. Inculcate professional attitudes, leadership qualities and social responsibility. 6. Equip students with industry-relevant skills and experiences for successful careers. 7. Enhance critical thinking and innovative problem solving abilities.
Relevance of PSOs to the local, regional, national, and global developmental needs	<p>Provide In-depth Knowledge: The program aims to provide students with a comprehensive understanding of the key concepts, theories, and methodologies in Data Science. It covers a range of topics including Data Analysis and Visualization, Statistical Methods for Data Science, data visualization, and SPARK Technologies, Soft Computing ,etc. enabling students to develop a deep knowledge base in these areas.</p> <p>Develop Programming Skills: The program aims to equip students with strong programming skills by providing hands on experience with different tools and technologies. Students will gain proficiency in designing front-end and back-end solutions, enhancing their ability to develop robust and scalable applications.</p> <p>Foster Problem-solving Abilities: The program aims to enhance students' problem-solving abilities by training them to approach real-world data challenges critically and creatively. Students will learn to identify problems, design appropriate data analysis strategies, and develop innovative solutions using their knowledge of Computer Science.</p> <p>Encourage Collaboration and Teamwork: The program aims to foster collaboration and teamwork skills among students, recognizing that computer science projects often require interdisciplinary collaboration. Students will learn to effectively communicate, collaborate, and contribute as part of a team, preparing them for collaborative work environments.</p> <p>Foster Industry Relevance: The program aims to stay up-to-date with industry trends and technologies to ensure graduates are well-prepared for the demands of the Data Science job market. Through industry partnerships and internships, students will have the opportunity to gain practical</p>

experience and stay in touch with the latest advancements in the field.

Professional Development: The program aims to prepare students for successful careers in the field of Data Science. In addition to technical skills, students will develop professional skills such as teamwork, project management, and leadership. The program may also provide networking opportunities, internships, or collaborations with industry partners to enhance students' industry readiness and employability.

Cultivate Research Skills: The program aims to cultivate research skills among students by providing a Research Methodology Course and encouraging participation in research projects. Students will learn to conduct literature reviews, design experiments, analyze data, and present their findings, fostering a research oriented mindset and contributing to the advancement of Data Science.

**Master of Science (Data Science) Programme
Under Choice Based Credit System (CBCS)
Course Structure (Autonomous)
Department of Data Science**

(To be implemented from Academic Year 2026-27)

Course Code	Semester III	Credits	Course Code	Semester IV	Credits
	<i>Major Mandatory</i>			<i>Major Mandatory</i>	
26_SDSDT901	Deep Neural Networks	04	26_SDSDT001	Next Generation Databases	04
26_SDSDP902	Deep Neural Networks Practical	02	26_SDSDP002	Next Generation Databases Practical	02
26_SDSDT903	Text Mining and Natural Language Processing	04	26_SDSDT003	Optimization Methods for Data Science	04
26_SDSDP904	Text Mining and Natural Language Processing Practical	02	26_SDSDP004	Optimization Methods for Data Science Practical	02
26_SDSDT905	Data Compliance and Auditing	02			
	<i>Major Electives (Any 1)</i>			<i>Major Electives (Any 1)</i>	
26_SDSET906	Image and Video Analytics	02	26_SDSET005	Blockchain Technologies for Data Science	02
26_SDSEP907	Image and Video Analytics Practical	02	26_SDSEP006	Blockchain Technologies for Data Science Practical	02
OR			OR		
26_SDSET908	Econometrics and Finance	02	26_SDSET007	Financial Risk Analytics and Management	02
26_SDSEP909	Econometrics and Finance Practical	02	26_SDSEP008	Financial Risk Analytics and Management Practical	02
OR			OR		
26_SDSET910	Data Science for Agriculture	02	26_SDSET009	Legal Analytics	02
26_SDSEP911	Data Science for Agriculture Practical	02	26_SDSEP010	Legal Analytics Practical	02
	<i>Research Project (RP)</i>			<i>Research Project (RP)</i>	
26_SDSPE912	Research Project	04	26_SDSPE011	Research Project	06
Total Credits		22	Total Credits		22

**Syllabus of Courses of M.Sc. Data Science
Programme at Semester III
with Effect from the Academic Year 2026-2027**

Nomenclature of the Course	Deep Neural Networks	
Class	M.Sc. Data Science	
Semester	III	
Course Code	26_SDSDT901	
No. of Credits	4	
Nature	Theory	
Type	Major Mandatory	
Highlight revision specific to employability/ entrepreneurship/ skill development	The Deep Neural Networks course enhances employability, entrepreneurship, and practical AI skills. It covers ML and DL fundamentals, CNNs, RNNs, GANs, Transformers, and BERT, focusing on architectures, training, optimization, evaluation, and real-world applications in computer vision, NLP, and time-series. The course equips learners with industry-ready skills, encourages AI experimentation, and fosters innovation for data-driven solutions.	
Course Outcomes:		
The learner will be able to :		
CO1: Understand the basics of Machine Learning and Deep Learning, neural network architectures, CNNs, RNNs, GANs, and Transformers.		
CO2: Implement deep learning models for image, sequence, and text-based applications using CNNs, RNNs, and attention-based architectures.		
CO3: Apply optimization techniques, regularization, and evaluation metrics to train, evaluate, and improve deep learning models.		
CO4: Explore and adapt recent advances in deep learning, generative models, and reinforcement learning to solve practical and research problems.		
Syllabus:		
Unit No.	Unit Title	Sub Titles (Learning Points)
1	Introduction to Machine Learning and Deep Learning	Basics of Machine Learning: types - supervised, unsupervised, reinforcement, difference between ML and DL, Neural Network architecture -input, hidden, output layers, activation functions -ReLU, Sigmoid, training process -forward and backward propagation, Optimization Algorithms -Gradient Descent, Adam, RMSProp, Overfitting, Underfitting, Regularization techniques Dropout, Early Stopping, L1/L2, Evaluation Metrics -Accuracy, Precision, Recall, F1-score, challenges in deep learning -vanishing gradients, computational cost.

2	Convolutional Neural Networks (CNNs)	Introduction to CNNs, Architectural elements -convolution, pooling, fully connected layers-, convolution and pooling operations, feature extraction and pattern recognition, activation functions in CNNs, Advanced CNN architectures: AlexNet, VGG, GoogLeNet, ResNet, applications of CNNs in image classification, face recognition, and object detection.
3	Recurrent Neural Networks (RNNs)	Basics of RNNs, architecture and working, sequence data processing, Backpropagation Through Time (BPTT), challenges -vanishing and exploding gradients, Advanced RNN structures: LSTM (Long Short-Term Memory), GRU (Gated Recurrent Units), applications of RNNs in text generation, speech recognition, and time-series prediction.
4	Generative Models and Reinforcement Learning	Generative Adversarial Networks (GANs): architecture, working of generator and discriminator, training process, Conditional GANs (cGAN), applications in image generation and enhancement, Transformers: architecture, self-attention mechanism, positional encoding, BERT (Bidirectional Encoder Representations from Transformers): architecture, training, fine-tuning, applications in NLP, Deep Reinforcement Learning: concept, Deep Q-Network (DQN), applications in gaming, robotics, and automation.

Reference Book:

1. Deep Learning by Ian Goodfellow, Yoshua Bengio, and Aaron Courville, MIT Press, 2016
2. Neural Networks and Deep Learning: A Textbook by Charu C. Aggarwal, Springer, 2018
3. Deep Learning with Python by François Chollet, Manning Publications, 2017
4. Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play by David Foster, O'Reilly Media, 2019
5. Transformers for Natural Language Processing by Denis Rothman, Packt Publishing, 2021
6. Deep Reinforcement Learning Hands-On by Maxim Lapan, Packt Publishing, 2018

Teaching Plan:

Unit No.	Unit Title	Teaching Methods	No. of Lectures
1	Introduction to Machine Learning and Deep Learning	Classroom Teaching and ICT	15
2	Convolutional Neural Networks (CNNs)	Classroom Teaching and ICT	15
3	Recurrent Neural Networks (RNNs)	Classroom Teaching and ICT	15
4	Generative Models and Reinforcement Learning	Classroom Teaching and ICT	15

**Syllabus of Courses of M.Sc. Data Science
Programme at Semester III
with Effect from the Academic Year 2026-2027**

Nomenclature of the Course	Deep Neural Networks Practical
Class	M.Sc. Data Science
Semester	III
Course Code	26_SDSDT902
No. of Credits	2
Nature	Practical
Type	Major Mandatory
Highlight revision specific to employability/ entrepreneurship/ skill development	The Deep Neural Networks Practicals course is designed to enhance employability, entrepreneurship, and practical AI skills. Students gain hands-on experience with TensorFlow, Keras, and PyTorch, implementing models such as CNNs, RNNs, LSTMs, GRUs, Autoencoders, GANs, and Transformers. Real-world applications include image classification, time series prediction, sequence generation, sentiment analysis, machine translation, and reinforcement learning.
Course Outcomes:	
The learner will be able to :	
CO1: Understand the architectures and workings of various deep learning models.	
CO2: Implement deep learning solutions for complex problems in computer vision, natural language processing.	
CO3: Evaluate and optimize deep learning models using modern techniques and frameworks.	
CO4: Adapt with cutting-edge research in deep learning to formulate and solve novel research questions.	
Note: The following set of practical's could be implemented in Python, TensorFlow, Keras, PyTorch, Jupyter/Colab, and GPU support.	
Syllabus:	
Sr. No.	Aim of the Practical
1	Implement a Basic Neural Network in TensorFlow.
2	Implement CNN with Keras for image classification.
3	Implement ResNet in PyTorch.
4	Implement RNN for Time Series Prediction.

5	Implement LSTM for Sequence Generation.
6	Implement GRU-based model for Sentiment Analysis.
7	Implement Text Classification Using Simple RNN (Not LSTM/GRU).
8	Build a Character-Level RNN for Name or Word Generation.
9	Apply Transformer Model for Machine translation task.
10	Apply BERT for Text Classification.
11	Implement a GAN to generate images.
12	Implement Early Stopping on a Classification Task.

Text Books:

1. Deep Learning by Ian Goodfellow, Yoshua Bengio, and Aaron Courville, MIT Press, 2016
2. Neural Networks and Deep Learning: A Textbook by Charu C. Aggarwal, Springer, 2018

Reference Books:

1. Pattern Recognition and Machine Learning by Christopher M. Bishop, Springer, 2006
2. Machine Learning: A Probabilistic Perspective by Kevin P. Murphy, MIT Press, 2012
3. Deep Learning for Computer Vision by Rajalingappaa Shanmugamani, Packt Publishing, 2018
4. Transformers for Natural Language Processing by Denis Rothman, O'Reilly Media, 2021

Teaching Plan:

Practicals	Unit Title	Teaching Methods	No. of Lectures
All	All Practical's	Lab Session with ICT	60

Syllabus of Courses of M.Sc. Data Science
Programme at Semester III
with Effect from the Academic Year 2026-2027

Nomenclature of the Course	Text Mining and Natural Language Processing	
Class	M.Sc. Data Science	
Semester	III	
Course Code	26_SDSDT903	
No. of Credits	4	
Nature	Theory	
Type	Major Mandatory	
Highlight revision specific to employability/ entrepreneurship/ skill development	Strengthen Python-based NLP coding, text preprocessing, feature engineering, and model building skills. Gain hands-on experience with RNNs, LSTMs, Transformers, and LLMs for real-world applications. Emphasize project implementation, sentiment analysis, topic modeling, and text classification to enhance employability, entrepreneurship, and practical AI/ML skill development.	
Course Outcomes:		
The learner will be able to :		
CO1: Able to analyze NLP theory and challenges, and apply NLTK/spaCy for fundamental text processing.		
CO2: Execute text pre-processing techniques and differentiate between BoW, TF-IDF, and modern vector embeddings.		
CO3: Implement classification models and apply core NLP methods (topic modeling, sentiment analysis, and named entity recognition).		
CO4: Implement and fine-tune deep neural network and Transformer models for complex NLP applications.		
Syllabus:		
Unit No.	Unit Title	Sub Titles (Learning Points)
1	Introduction to Compiler, Text Mining and NLP	Introduction to Compiler: Definition, purpose, and phases of a compiler,Compilation vs interpretation,Lexical analysis: tokens, lexemes,Parsing: syntax trees, grammars,Semantic analysis and code generation (overview)

		<p>Overview of text mining and NLP: Definition and scope, Differences between text mining, NLP, and information retrieval,</p> <p>Applications in real-world domains</p> <p>Applications of NLP in various domains: Chatbots, sentiment analysis, spam detection, Recommendation systems, Case studies in various industries</p> <p>Key challenges in text mining: (e.g., handling unstructured data, language ambiguity)</p> <p>Introduction to Python libraries for NLP: NLTK, spaCy)</p>
2	Text Pre-processing and Representation, Parsing & POS Tagging	<p>Introduction to Text Pre-processing: Importance of preprocessing in NLP, Overview of the NLP pipeline. Tokenization: Word and sentence tokenization, Handling punctuation and special characters</p> <p>Difference between stemming and lemmatization, Algorithms: Porter Stemmer, Snowball Stemmer, WordNet Lemmatizer, Hands-on exercises. Stopwords and their impact, Lowercasing, case normalization, Handling accents, special characters. Handling misspellings, Removing URLs, emojis, HTML tags, and punctuation, Example preprocessing pipelines. Concept of BoW, Creating BoW vectors in Python, Sparse vs dense representations</p> <p>Term Frequency, Inverse Document Frequency, TF-IDF vectorization, Comparison with BoW. Word Embeddings: Introduction to embeddings, Word2Vec, GloVe, FastText basics. Sentence and Document Embeddings: Sentence embeddings (Doc2Vec, SBERT basics), Document-level representations. Part-of-Speech (POS) Tagging: Definition and importance, POS tagging rules vs statistical models. Parsing: Constituency parsing, Dependency parsing, Introduction to parse trees. Advanced Parsing Techniques: Practical applications of parsing, Parsing in English and multilingual texts. Mini Project / Case Study.</p>
3	Text Classification, Topic Modeling & Sentiment Analysis	<p>Introduction to Text Classification, Data Preparation for Classification: Preprocessing recap, feature extraction (BoW, TF-IDF, embeddings). Naive Bayes for Text Classification: Multinomial & Bernoulli models, advantages/limitations. Support Vector Machines (SVM): Linear & kernel SVM for text. Logistic Regression: Multiclass logistic regression, softmax. Model Evaluation Metrics: Accuracy, precision, recall, F1-score, confusion matrix, ROC-AUC. Advanced Classification Techniques: RNN, LSTM, Transformer-based classification (BERT/DistilBERT). Sentiment Analysis:</p>

		Lexicon-based vs ML-based approaches, applications. Topic Modeling Introduction: LDA, concept, use cases.
4	Advanced NLP, Transformers & LLMs	Introduction to Advanced NLP: Limitations of classical NLP, motivation for deep learning. Recurrent Neural Networks (RNNs): Architecture, sequence modeling, NLP applications. Long Short-Term Memory (LSTM) Networks: LSTM architecture, gates, handling long-term dependencies. Sequence-to-Sequence Models: Encoder-decoder architecture, translation, summarization. Attention Mechanism: Motivation, multi-head attention, improvement over RNNs. Introduction to Transformers: Encoder-decoder, multi-head attention, advantages over RNN/LSTM. BERT: Bidirectional Transformers: Pre-training tasks, fine-tuning for downstream tasks. GPT and Generative Transformers: Autoregressive modeling, text generation, summarization. Large Language Models (LLMs): GPT-3, GPT-4, LLaMA, ChatGPT; capabilities & architecture. LLM Applications & Prompt Engineering. Ethical Considerations & Challenges,

Prescribed Text/s (If any):

1. Speech and Language Processing by Daniel Jurafsky and James H. Martin, 2008, Pearson Prentice Hall
2. Natural Language Processing with Python by Steven Bird, Ewan Klein, and Edward Loper, 2009, O'Reilly Media
3. Neural Network Methods for Natural Language Processing by Yoav Goldberg, 2017, Morgan & Claypool Publishers

Teaching Plan:

Unit No.	Unit Title	Teaching Methods	No. of Lectures
1	Introduction to Compiler, Text Mining and NLP	Classroom Teaching and ICT	15
2	Text Pre-processing and Representation, Parsing & POS Tagging	Classroom Teaching and ICT	15
3	Text Classification, Topic Modeling & Sentiment Analysis	Classroom Teaching and ICT	15
4	Advanced NLP Techniques and Transformer Models	Classroom Teaching and ICT	15

Syllabus of Courses of M.Sc. Data Science
Programme at Semester III
with Effect from the Academic Year 2026-27

Nomenclature of the Course	Text Mining and Natural Language Processing Practical
Class	M.Sc. Data Science
Semester	III
Course Code	26_SDSDP904
No. of Credits	4
Nature	Practical
Type	Major Mandatory
Highlight revision specific to employability/ entrepreneurship/ skill development	Mastery of this full NLP toolkit—from advanced normalization and feature engineering (TF-IDF, N-grams) to building functional systems for Sentiment Analysis, NER, and Text Classification—drives high employability. These hands-on skills prepare you for immediate placement as an ML Engineer or Data Scientist. Crucially, this practical ability to create deployable solutions enables entrepreneurs to launch innovative, automated text processing and business intelligence products.

Course Outcomes:

The learner will be able to :

CO1: Prepare raw text using tokenization, stop-word removal, and correctly apply/compare stemming and lemmatization techniques.

CO2: Implement Part-of-Speech (POS) tagging and N-gram models for syntactic analysis and foundational feature engineering.

CO3: Quantify term importance using the TF-IDF method for machine learning tasks.

CO4: Develop and evaluate systems for key information extraction tasks, including Named Entity Recognition (NER) and extractive summarization.

CO5: Build and implement end-to-end NLP pipelines for practical text classification and sentiment analysis tasks.

The following practical's will be implemented using Python/Google Colab

Syllabus:

Sr. No.	Aim of the Practical
1	Implement tokenization, stemming, lemmatization, stopword removal, and text normalization.
2	Convert text into BoW, TF-IDF, and word/document embeddings for NLP tasks.
3	Apply Part-of-Speech tagging on sample corpora and analyze syntactic structure.
4	Identify and extract entities (persons, organizations, locations) from text.

5	Build and evaluate Naive Bayes, SVM, and Logistic Regression classifiers.
6	Implement RNN/LSTM-based classifiers for sentiment analysis or topic prediction.
7	Perform Latent Dirichlet Allocation (LDA) on a corpus and interpret discovered topics.
8	Analyze opinions from reviews or social media using ML and deep learning approaches.
9	Apply seq2seq models with attention for text summarization or translation tasks.
10	Fine-tune BERT or GPT models for classification, NER, or other downstream tasks.
11	Explore prompt engineering and query-based tasks using GPT-3/GPT-4 or similar models.
12	Design and deploy a rule-based or AI-based chatbot for a practical domain (customer support, FAQ, or simple conversation).

Teaching Plan:

Unit No.	Unit Title	Teaching Methods	No. of Lectures
All	All Practicals	Lab Session with ICT	60

**Syllabus of Courses of M.Sc. Data Science
Programme at Semester III
with Effect from the Academic Year 2026-27**

Nomenclature of the Course	Data Compliance and Auditing	
Class	M.Sc. Data Science	
Semester	III	
Course Code	26_SDSDT905	
No. of Credits	2	
Nature	Theory	
Type	Major Mandatory	
Highlight revision specific to employability/ entrepreneurship/ skill development	The course aims to develop practical skills in data auditing, compliance documentation, foster entrepreneurship by encouraging learners to explore opportunities in data privacy consulting and audit services, promoting innovative thinking and independent professional practice. Additionally, the course is designed to build employability by providing hands-on exposure to industry-relevant compliance tools, frameworks, and methodologies used in contemporary data governance environments.	
Course Outcomes:		
The learner will be able to :		
CO1: Gain a comprehensive understanding of major data protection and privacy regulations such as GDPR, and their implications for businesses.		
CO2: Develop the ability to plan, conduct, and document data audits using industry-standard methodologies and tools to ensure data accuracy and compliance.		
Syllabus:		
Unit No.	Unit Title	Sub Titles (Learning Points)
1	Data Compliance	General Data Protection Regulation (GDPR) - Data Protection, Material scope of the GDPR, The building blocks of the GDPR, Compliance framework – the standards of protection, Eprivacy- Regulating the electronic communications sector, The relationship between data protection and ePrivacy, Operational Data Protection - The three layers of an organisation, implementing data protection in the people layer, implementing data protection in the paper layer. Core Law of Data Protection- The principles of data protection, Lawfulness, fairness and transparency – the first principle, Purpose limitation – the second principle, Data minimisation – the third principle.

		The Rights of Data Subjects- Transparency and information rights, Rights over data processing. Mechanisms to Support Operational Compliance. Programmatic Approaches for Delivering Data Protection By Design And Default.
2	Data Auditing	<p>Information Security Performance Metrics and Audit: Pre-audit checklist, Information Gathering, Vulnerability Analysis, External Security Audit, Internal Network Security Audit, Firewall Security Audit, IDS Security Auditing, Social Engineering Security Auditing, Web Application Security Auditing.</p> <p>Vulnerability Management: Information Security Vulnerabilities – Threats and Vulnerabilities, Human-based Social Engineering, Computer-based Social Engineering, Social Media Countermeasures, Vulnerability Management – Vulnerability Scanning, Testing, Threat management, Remediation etc.</p> <p>Information Security Assessments: Vulnerability Assessment, Classification, Types of Vulnerability Assessment, Vulnerability Assessment Phases, Vulnerability Analysis Stages, Characteristics of a Good Vulnerability Assessment Solutions & Considerations, Vulnerability Assessment Reports – Tools and choosing a right Tool, Information</p> <p>Security Risk Assessment, Risk Treatment, Residual Risk, Risk Acceptance, Risk Management Feedback Loops etc.</p> <p>Configuration Reviews: Introduction to Configuration Management, Configuration Management Requirements-Plan Control, Development of configuration Control Policies, Testing Configuration Management etc.</p>

Prescribed Text/s (If any):

1. Data Protection and Compliance, 2nd Edition by Stewart Room, Michelle Maher, Niall O'Brien, Adam Panagiotopoulos, Shervin Nahid, Richard Hall, Tughan Thuraisingam, James Drury-Smith, Simon Davis, Mark Hendry, Jamie Taylor, Ben Johnson Released November 2021
2. Data auditing The Ultimate Step-By-Step Guide by Gerardus Blokdyk, 2018, 5STARCOOKS
3. Assessing Information Security (strategies, tactics, logic and framework) by A Vladimirov, K.Gavrilenko, and A.Michajlowski, Addison-Wesley Professional , 2010

Teaching Plan:

Unit No.	Unit Title	Teaching Methods	No. of Lectures
1	Data Compliance	Classroom Teaching and ICT	15
2	Data Auditing	Classroom Teaching and ICT	15

Syllabus of Courses of M.Sc. Data Science
Programme at Semester III
with Effect from the Academic Year 2026-2027

Nomenclature of the Course	Image and Video Analytics	
Class	M.Sc. Data Science	
Semester	III	
Course Code	26_SDSET906	
No. of Credits	2	
Nature	Theory	
Type	Major Electives	
Highlight revision specific to employability/ entrepreneurship/skill development	The Image and Video Analytics course enhances employability and entrepreneurial potential by developing advanced skills in computer vision, image processing, and AI-based analytics. Students gain hands-on experience with industry tools such as OpenCV, TensorFlow, and PyTorch to design real-world solutions for sectors like security, healthcare, media, and smart cities. The course encourages innovation and entrepreneurship by guiding learners to create AI-driven visual intelligence products and startup ideas.	
Course Outcomes:		
The learner will be able to :		
CO1: Explain fundamental concepts of digital images including pixels, resolution, aspect ratio, color models, and formats.		
CO2: Apply preprocessing operations such as resizing, filtering, noise removal, and thresholding to improve image quality for further analysis.		
Syllabus:		
Unit No.	Unit Title	Sub Titles (Learning Points)
1	Image Analytics Fundamentals	<p>Digital Image Basics What is an image? Pixel, resolution, aspect ratio, Color models: RGB, HSV, Grayscale, CMYK (basic use cases), Image formats: JPEG, PNG, TIFF, BMP – compression and quality trade-offs.</p> <p>Image Preprocessing Image resizing, cropping, scaling, Filtering techniques: blur, sharpening, median filter, Gaussian filter, Noise in images: types (salt & pepper, Gaussian noise) and removal methods, Thresholding and binarization (Otsu's method, adaptive thresholding) Feature Extraction, Edge detection: Sobel, Prewitt, Canny Corner and blob detection (Harris corner, SIFT, SURF – only concepts), Texture analysis (histogram, GLCM) Color histograms for object recognition.</p> <p>Applications of Image Analytics</p>

		Image classification basics, Face detection using Haar cascades, OCR (Optical Character Recognition) – extracting text from images.
2	Video Analytics & Machine Learning Applications	<p>Video Representation & Motion Analysis Frames, frame rate (FPS), resolution, Video compression and codecs (MP4, AVI, H.264 basics), Difference between still image and video data, Motion detection: temporal changes, frame differencing, Background subtraction (Gaussian Mixture Models, simple running average)</p> <p>Object Detection & Tracking in Video ROI (Region of Interest) selection, Object tracking: Centroid tracking, KLT tracker, Mean-Shift, CamShift Multiple object tracking challenges.</p> <p>Machine Learning & Deep Learning for Vision Neural Networks basics for vision, CNNs: filters, pooling, activation functions, Transfer learning: VGG, ResNet, MobileNet,</p> <p>Practical Applications Healthcare: medical imaging (X-rays, MRI, CT scans), Security: face recognition, crowd monitoring, Retail: customer behavior, shelf monitoring, Entertainment: video recommendation, AR/VR, Smart surveillance, traffic monitoring, sports analytics, retail analytics.</p> <p>Ethical & Social Aspects Privacy concerns in surveillance ,Bias in datasets and AI models, Responsible use of AI in vision.</p>

Prescribed Text/s (If any):

1. Digital Image Processing – Rafael C. Gonzalez, Richard E. Woods 4th Edition, 2018, Pearson
2. Fundamentals of Digital Image Processing – Anil K. Jain 1st Edition, 1989, Prentice Hall
3. Computer Vision: Algorithms and Applications – Richard Szeliski 2nd Edition, 2022, Springer (1st Edition was 2010, free PDF available from author’s site).
4. Learning OpenCV 4: Computer Vision with Python and C++ – Adrian Kaehler, ary Bradski, 3rd Edition, 2019, O’Reilly Media
5. Computer Vision: Models, Learning, and Inference – Simon J.D. Prince 1st Edition, 2012, Cambridge University Press

Teaching Plan:

Unit No.	Unit Title	Teaching Methods	No.of Lectures
1	Image Analytics Fundamentals	Classroom Teaching and ICT	15
2	Video Analytics & Machine Learning Applications	Classroom Teaching and ICT	15

Syllabus of Courses of M.Sc. Data Science
Programme at Semester III
with Effect from the Academic Year 2026-2027

Nomenclature of the Course	Image and Video Analytics Practical
Class	M.Sc. Data Science
Semester	III
Course Code	26_SDSEP907
No. of Credits	2
Nature	Practical
Type	Major Electives
Highlight revision specific to employability/ entrepreneurship/skill development	This practical equips students with foundational and advanced image processing techniques essential for careers in Image and Video Analytics. Learners gain hands-on experience in image representation, format conversion, enhancement, noise handling, and filtering. Key skills include thresholding, edge detection, and feature extraction—crucial for computer vision, surveillance, and media industries. Mastery of these concepts prepares students for roles in AI-driven visual systems, quality inspection, and video content analysis.

Course Outcomes:

The learner will be able to :

CO1: Understand Digital Images and Preprocessing Techniques.

CO2: Feature Extraction and Image Analysis.

CO3: Understand applications of Image Analytics.

CO4: Understand Machine Learning and Deep Learning for Vision.

The following practical's will be implemented using Python

Syllabus:

Sr.No.	Aim of the Practical
1	Display an image, print its pixel values, resolution, and aspect ratio.
2	Convert an image between RGB, HSV, and Grayscale formats and observe the difference.
3	Save an image in different formats (JPEG, PNG, BMP, TIFF) and compare file sizes and quality.
4	Perform resizing, cropping, and scaling operations on an image.
5	Apply blur, Gaussian, median, and sharpening filters to enhance an image.
6	Add Salt & Pepper and Gaussian noise to an image and remove it using filters.
7	Apply simple, Otsu's, and adaptive thresholding to convert images into binary.
8	Detect edges using Sobel, Prewitt, and Canny operators; visualize Harris corners.
9	Create and plot image histograms (color & grayscale), and calculate texture features using GLCM.
10	Detect faces using Haar cascades and extract text using Tesseract OCR.
11	Read a video, extract frames, and detect motion using frame differencing or background subtraction.

12	Track a moving object using Mean-Shift or CamShift algorithms.		
Prescribed Text/s (If any):			
<ol style="list-style-type: none"> 1. Digital Image Processing and Analysis: Computer Vision and Image Analysis, Scott E. Umbaugh, 4th Edition (2023), CRC Press 2. Digital Image Processing: An Algorithmic Introduction, Wilhelm Burger & Mark J. Burge, 3rd Edition (2021), Springer 3. Computer Vision: Algorithms and Applications, Richard Szeliski, 2nd Edition (2023), Springer 4. Digital Image Processing and Analysis: Computer Vision and Image Analysis, Scott E. Umbaugh, 4th Edition (2023), CRC Press 5. Computer Vision: Algorithms and Applications, Richard Szeliski, 2nd Edition (2023), Springer 			
Teaching Plan:			
Practicals	Unit Title	Teaching Methods	No. of Lectures
All	All Practical's	Lab Session with ICT	60

Syllabus of Courses of M.Sc. Data Science
Programme at Semester III
with Effect from the Academic Year 2026-2027

Nomenclature of the Course	Econometrics and Finance	
Class	M.Sc. Data Science	
Semester	III	
Course Code	26_SDSET908	
No. of Credits	2	
Nature	Theory	
Type	Major Electives	
Highlight revision specific to employability/ entrepreneurship/ skill development	The course strengthens employability by developing analytical and financial modeling skills essential for data-driven roles in business and finance. It promotes entrepreneurship and skill development through practical training in econometric analysis and spreadsheet-based financial decision-making.	
Course Outcomes:		
The learner will be able to :		
CO1: Students will be able to apply econometric models to analyse real-world data, test assumptions, and interpret regression results for informed decision-making.		
CO2: Students will be able to design, test, and validate financial models using spreadsheets for effective business and financial decision-making.		
Syllabus:		
Unit No.	Unit Title	Sub Titles (Learning Points)
1	Econometrics	Introduction to Econometrics– definitions, scope, methodology, types. Uni-variate regression model, Assumptions, Gauss Markov Theorem. Multivariate regression model and Inferential Analysis. Test of significance of parameter estimates- Test of goodness of fit and correlation Problem and Methods to detect Heteroscedasticity, Multicollinearity and Autocorrelation in the Regression Model, its causes, consequences and Remedial Measures. Dummy variables- Nature of Dummy variables, Use of Dummy Variables, Errors in Variables and its consequences.
2	Financial Modelling	Introduction to financial modelling-objectives of financial modelling, spreadsheet features, techniques, best practices in spreadsheet design. Designing models - Model Design and structure, Building business case models, spreadsheet techniques and methods. Auditing and Testing-Essential testing and auditing

		techniques, Testing financial analysis model with cash flows and ratios, Debugging and checking a financial model. Macros and Security-Writing and auditing and macros, Spreadsheet security. Model Completion: Model review; Documentation; Final audit.
--	--	---

Prescribed Text/s (If any):

1. Damodar N. Gujarathi: Basic Econometrics, New Delhi: Tata McGraw Hill, 5th edition, 2009
2. J. Johnston: Econometric Methods, McGraw Hill, 4th edition, 1997
3. John Hull: Futures, Options and Other Derivatives, Prentice Hall, 10th edition, 2021
4. Financial Valuation and Modeling by Sheeba Kapil, Wiely, 3rd Edition, 2021

Teaching Plan:

Unit No.	Unit Title	Teaching Methods	No. of Lectures
1	Econometrics	Classroom Teaching and ICT	10
2	Financial Modelling	Classroom Teaching and ICT	10

**Syllabus of Courses of M.Sc. Data Science
Programme at Semester III
with Effect from the Academic Year 2026-2027**

Nomenclature of the Course	Econometrics and Finance Practical
Class	M.Sc. Data Science
Semester	III
Course Code	26_SDSEP909
No. of Credits	2
Nature	Practical
Type	Major Electives
Highlight revision specific to employability/ entrepreneurship/skill development	The course enhances employability by equipping learners with practical skills in econometric modeling, financial analysis, and spreadsheet-based decision-making. It fosters entrepreneurship and skill development through hands-on application of regression techniques, model design, automation, and data-driven problem-solving.
Course Outcomes:	
The learner will be able to :	
CO1: Apply regression models (univariate & multivariate) and diagnose econometric problems using appropriate detection methods and remedies.	
CO2: Interpret the role of dummy variables in regression analysis for better model specification.	
CO3: Design and implement financial models in spreadsheets following best practices in structure and auditing.	
CO4: Develop secure, efficient spreadsheet models using macros, testing, and documentation techniques.	
Note: Working on - Microsoft Excel / R Programming/ Python	
Syllabus:	
Sr. No.	Aim of the Practical
1	Perform a Univariate Regression Model to estimate the relationship between a dependent and a single independent variable and interpret the results.
2	Perform a Multivariate Regression Model to analyze the relationship between a dependent variable and multiple independent variables, and interpret the parameter estimates and model fit.
3	Apply suitable econometric methods to detect heteroscedasticity in a regression model and implement appropriate remedial measures to correct it.
4	Detect multicollinearity in a regression model through suitable methods (e.g., correlation matrix, VIF), and apply appropriate remedial measures.
5	Detect the presence of autocorrelation in a regression model with appropriate tests (e.g., Durbin-Watson), and apply suitable remedial measures if required.
6	Create and include dummy variables in a regression model and interpret their impact on the dependent variable.

7	Design a financial model in a spreadsheet applying best practices in structure, formatting, and organization to ensure clarity, accuracy, and usability.		
8	Develop a financial or econometric model in a spreadsheet by planning its design and structure, clearly defining inputs, calculations, and outputs for effective analysis and decision-making.		
9	Perform auditing and testing of a financial or econometric model in a spreadsheet to verify accuracy, identify errors, and ensure the reliability of calculations and outputs.(Use MS Excel)		
10	Create and audit macros in a spreadsheet model, and implement security measures to protect data integrity and restrict unauthorized access.		
11	Perform Model Validation and Forecasting using Regression Models.		
12	Perform Sensitivity and Scenario Analysis in a Financial or Econometric Model.		
Prescribed Text/s (If any):			
Teaching Plan:			
Practicals	Unit Title	Teaching Methods	No. of Lectures
All	All Practical's	Lab Session with ICT	60

Syllabus of Courses of M.Sc. Data Science
Programme at Semester III
with Effect from the Academic Year 2026-2027

Nomenclature of the Course	Data Science for Agriculture	
Class	M.Sc. Data Science	
Semester	III	
Course Code	26_SDSET910	
No. of Credits	2	
Nature	Theory	
Type	Major Elective	
Highlight revision specific to employability/ entrepreneurship/ skill development	The course develops hands-on skills in agricultural data collection, preprocessing, and predictive modeling to enhance employability. Students learn remote sensing, GIS, and precision agriculture techniques for industry-relevant applications. Practical projects on smart farming solutions, yield prediction dashboards, and agritech startups foster entrepreneurship. Training in tools like Python, Scikit-learn, QGIS, and IoT sensor data analysis builds technical expertise. The curriculum emphasizes data-driven decision-making, problem-solving, and critical thinking to strengthen overall skill development.	
Course Outcomes:		
The learner will be able to:		
CO1: Demonstrate proficiency in agricultural data science by collecting, preprocessing, and analyzing datasets using exploratory and predictive modeling techniques to derive actionable insights.		
CO2: Apply remote sensing, GIS, and precision agriculture methods to design sustainable, data-driven solutions that optimize resources and minimize environmental impacts.		
Syllabus:		
Unit No.	Unit Title	Sub Titles (Learning Points)
1	Foundations of Agricultural Data Science	Introduction to the role of data science in agriculture: Key challenges and opportunities. Overview of agricultural data sources, Types of data in agriculture (soil data, weather data, crop data, etc.), Importance of big data in agriculture, Data Collection Techniques: Remote sensing and satellite imagery, IoT and sensor networks in agriculture, Manual data collection methods. Data Management: Cloud storage solutions for agricultural data. Data Cleaning and Preprocessing. Exploratory data Analysis (EDA): Techniques for summarizing and visualizing data, identifying patterns and correlations, Using EDA tools and libraries

2	Intelligent Agricultural Systems through Machine Learning and AI	<p>Machine Learning Fundamentals: Introduction to supervised and unsupervised learning, Common algorithms (regression, classification, clustering), Model evaluation and validation.</p> <p>Predictive Analytics in Agriculture: Yield prediction models, Disease and pest prediction, Weather forecasting and its impact on agriculture.</p> <p>Advanced Machine Learning Techniques: Deep learning applications in agriculture, Computer vision for crop monitoring, Natural language processing for agricultural data, Applications of machine learning in precision farming.</p>	
Prescribed Text/s (If any):			
<ol style="list-style-type: none"> 1. Data Science for Agriculture by Simon Stelling, April 14, 2020 2. Data Science for Agriculture" by David L. Bunge 3. Practical Data Science for Agriculture" by Kellee Koenig 			
Teaching Plan:			
Unit No.	Unit Title	Teaching Methods	No. of Lectures
1	Foundations of Agricultural Data Science	Classroom Teaching and ICT	15
2	Intelligent Agricultural Systems through Machine Learning and AI	Classroom Teaching and ICT	15

**Syllabus of Courses of M.Sc. Data Science
Programme at Semester III
with Effect from the Academic Year 2026-2027**

Nomenclature of the Course	Data Science for Agriculture Practical
Class	M.Sc. Data Science
Semester	III
Course Code	26_SDSEP911
No. of Credits	2
Nature	Practical
Type	Major Elective
Highlight revision specific to employability/ entrepreneurship/ skill development	Practical sessions develop hands-on skills in data collection, preprocessing, and exploratory analysis for real-world agricultural datasets. Students gain experience with remote sensing, GIS, IoT sensors, and predictive modeling tools used in industry. Projects like crop yield prediction, irrigation optimization, and plant disease detection encourage entrepreneurial thinking. Hands-on training in Python, Scikit-learn, QGIS, OpenCV, and dashboard creation enhances technical competencies. Practical exercises strengthen problem-solving, data-driven decision-making, and critical analysis, improving employability and professional readiness.
Course Outcomes:	
The learner will be able to: CO1: Demonstrate proficiency in collecting, managing, and preprocessing agricultural data from multiple sources for analysis and decision support. CO2: Perform Exploratory Data Analysis (EDA) to identify trends, correlations, and patterns within agricultural datasets. CO3: Develop and apply predictive and machine learning models, along with remote sensing and GIS techniques, to support precision agriculture and resource optimization. CO4: Analyze and forecast market trends, supply chains, and agricultural productivity to enhance sustainability and profitability in farming systems.	
Note: Softwares : Python 3.0 above, Tools : Scikit-learn, XGBoost, OpenCV, GDAL, QGIS	
Use of Public dataset	
Syllabus:	
Sr. No.	Aim of the Practical
1	Collect agricultural data from web sources using Python-based web scraping tools, perform data cleaning and preprocessing to handle missing values and outliers, and conduct exploratory data analysis (EDA) to identify patterns, trends, and correlations in crop yield data.
2	Predict future weather patterns using historical weather data.
3	Build a machine learning model to predict crop yields based on various factors (soil type, weather, etc.).
4	Use satellite imagery to assess crop health and identify areas needing attention.

5	Classify land use patterns using remote sensing data and machine learning algorithms.		
6	Predict soil moisture levels using data from soil sensors and environmental factors		
7	Analyze irrigation data to optimize water usage and improve crop yield.		
8	Detect plant diseases using image processing techniques on leaf images		
9	Generate yield maps to visualize spatial variability in crop production.		
10	Predict market prices for agricultural products using historical price data and other relevant factors		
11	<p>A district agricultural research center wants to build a yield prediction model for rice. They have collected the following data for the last 10 years:</p> <p>Dataset Variables</p> <ol style="list-style-type: none"> 1. Rainfall (mm) – monthly 2. Average temperature (°C) – monthly 3. Soil nutrients: Nitrogen (N), Phosphorus (P), Potassium (K) 4. Soil pH 5. NDVI (Vegetation Index) 6. Fertilizer usage (kg/ha) 7. Historical rice yield (tons/ha) <p>Objective To detect which factors affect yield and to build a predictive model for next season.</p>		
12	<p>A farmer’s cooperative wants to implement an AI-based mobile application that identifies leaf diseases in tomato plants using images. They have collected:</p> <p>Dataset</p> <ul style="list-style-type: none"> ● 10,000 leaf images ● Three classes: <ol style="list-style-type: none"> 1. Healthy 2. Early Blight 3. Late Blight <p>Each image varies in lighting, camera type, and background.</p> <p>Objective Create a deep learning model (CNN) that can classify tomato leaf images with high accuracy.</p>		
Prescribed Text/s (If any):			
Teaching Plan:			
Practicals	Unit Title	Teaching Methods	No. of Lectures
All	All Practical’s	Lab Session with ICT	60

**Syllabus of Courses of M.Sc. Data Science
Programme at Semester III
with Effect from the Academic Year 2026-2027**

Nomenclature of the Course	RESEARCH PROJECT
Course Code	26_SDSPE912
Class	M.Sc. Data Science
Semester	III
No of Credits	04
Nature	Practical
Type	Research Project
Relevance with Employability/ Entrepreneurship/ Skill development	The research project outcomes for M.Sc DS students, encompassing research methodologies, statistical analysis, econometric techniques, and financial management strategies, are highly relevant for both employability and Entrepreneurship. Proficiency in these areas enhances analytical and problem-solving skills, making graduates valuable assets to employers in various fields. For entrepreneurs, these skills are crucial for informed decision making, market analysis, and effective business planning, contributing to the success and sustainability of their ventures in a dynamic business environment.

A student is expected to devote at least 2 to 3 months of effort to the Research Project. Students should submit a detailed research project report at the time of viva.

Course Outcomes

At the end of the Course, the Learner will be able to

1. Identify the research problem and formulate objectives.
2. Choose appropriate methodology with proper tools and techniques.
3. Analyze and interpret the data collected from different sources.
4. Make decision or find out conclusions on the basis of data analysis.
5. Recall and list key research paradigms and methodologies in the subject.
6. Explain the principles of statistical analysis and their application in subject of research.

Guidelines for Documentation of Research Project Proposal in Semester –III

Certified Spiral Bound Copy with Certificate is required to submit at the time of Viva Examination. A student should submit a Research Project Proposal report with the following details:

- **Title:** Title of the Research Project.
- **Objective:** A detailed objective of the proposal is needed.
- **Introduction/Background:**
- **Related works/Literature Survey:** A detailed survey of the relevant works done by others in the domain. The student is expected to refer to at least 20 recent (last five years) research papers in addition to textbooks and web links in the relevant topic.
- **Proposed Methodology:** - Describe the overall research design, including whether it will be quantitative, qualitative, or mixed-methods. Explain the rationale behind the chosen design and how it aligns with the research objectives. Explain the characteristics of the participants, including demographics, sample size, selection criteria, and recruitment methods. Outline the methods used for data collection, such as surveys, interviews, observations, or document analysis.
- **Research Challenges:** - List All research challenges related to data collection, design and implementation
- **Conclusion/ Summary:** - Write one-page review or summary of your research proposal with respect to above points.
- **References:** List of References

Certified Spiral Bound Copy with Certificate is required to submit at the time of Viva Examination.

Scheme of Examination: -

Total Credits	Total Hours	Marks	
		Internal	External
04	120	40	60

Internal Examination

A) Continuous Internal Evaluation:

Method		Marks	
Internal Viva 1		20	
Topic Weightage	Introduction	Objectives	Total
05	05	10	20
Internal Viva 2		20	
Proposed Methodology	Research Challenges & Conclusion	Documentation	Total
10	05	05	20

External Examination

A) External Evaluation:

Method				Marks			
External Viva				60			
Topic Weightage	Introduction	Objectives	Proposed Methodology	Research Challenges	Documents	Viva	Total
05	05	10	15	05	10	10	60

Format

1st page (Main Page)

Title of the problem of the Project

A Project Submitted

To

**R. P. Gogate College of Arts & Science and
R.V. Jogalekar College of Commerce (Autonomous), Ratnagiri**

of

Master in Commerce / Art/ Science

Under the Faculty of Commerce /Art/ Science

By

Name of Student

Under the Guidance of

Name of the Guiding Teacher

**R. P. Gogate college of Arts & Science and
R.V. Jogalekar College of Commerce (Autonomous),
Ratnagiri Advocate N.V. Joshi Road,
Near Ratnagiri District Court, Ratnagiri**

Month and Year

On separate page

Index

Chapter No	Title of the Chapter	Page No.
01		
02		
03		
04		
05		

On separate page

Declaration by learner

I the undersigned Miss/Mr. [Name of the learner] here by, declare that work embodied in this Research project work titled forms my own contribution to the research work carried out under the guidance of [Name of the guiding teacher] is a result of my own research work and has not been previously submitted to any other University or College for any other Degree/ Diploma.

Wherever reference has been made to previous works of others, it has been clearly indicated as such and included in the bibliography.

I, here by further declare that all information of this document has been obtained and presented in accordance with academic rules and ethical conduct.

Name and Signature of the learner

Certified by

Name and signature of the Guiding Teacher

On separate page

Acknowledgment

(Model structure of the acknowledgement)

To list who all have helped me is difficult because they are so numerous and the depth is so enormous.

I would like to acknowledge the following as being idealistic channels and fresh dimensions in the completion of this Research project.

I thank the **R. P. Gogate college of Arts & Science and R.V. Jogalekar College of Commerce (Autonomous), Ratnagiri** for giving me opportunity to do this project.

I would like to thank my **Principal, Prof. Dr M. R. Sakhalkar Sir**, for providing the necessary facilities required for completion of this project.

I take this opportunity to thank our Coordinator, for his moral support and guidance.

I would also like to express my sincere gratitude towards my project guide whose guidance and care made the project successful.

I would like to thank my College Library, for having provided various reference books and magazines related to my project.

Lastly, I would like to thank each and every person who directly or indirectly helped me in the completion of the project especially my Parents and Peers who supported me throughout my project.

Date:

Place:

(Head and BoS Chairman)

SEMESTER IV

Syllabus of Courses of M.Sc. Data Science
Programme at Semester IV
with Effect from the Academic Year 2026-2027

Nomenclature of the Course	Next Generation Databases
Class	M.Sc. Data Science
Semester	IV
Course Code	26_SDSDT001
No. of Credits	4
Nature	Theory
Type	Major Mandatory
Highlight revision specific to employability/ entrepreneurship/ skill development	The Next Generation Databases course is designed to boost employability, entrepreneurship, and practical skills. It provides hands-on training in SQL, NoSQL (MongoDB), Hadoop, Hive, HBase, Pig, and Neo4j, covering data modeling, querying, processing, and distributed systems. Students work on real-world projects, case studies, and cloud-based platforms, gaining job-ready database and analytics skills while fostering innovation and data-driven entrepreneurship.

Course Outcomes:

The learner will be able to :

CO1: Understand the fundamentals of Big Data, its applications in various domains and the technologies in the Bigdata ecosystem.

CO2: Understand the difference between Database system and data warehouse, the architecture and working of Enterprise Data warehouse, understand RDBMS model, perform SQL queries, importance of ACID properties.

CO3: Differentiate between ACID and BASE, Understand HDFS File Structure, Map Reduce Framework.

CO4: Understand the different types of NoSQL databases like document database, columnar database, and graph database.

Syllabus:

Unit No.	Unit Title	Sub Titles (Learning Points)
1	Big Data Foundations and Traditional Databases	Getting Started, Big Data, Facts About Big Data, Big Data Sources, Types of bigdata (structured, semi structured, unstructured), Usage of Big Data (Banking and Finance, Media, and communications etc), Data governance Policies and Procedures, Legacy Systems and Big Data, Data Storage, Data Processing, Big Data Technologies. Relational Database, Database Design Data Storage, Data warehouse and Data Mining, Information Retrieval, OLTP vs OLAP, Advantages of warehousing approach, Components of a Datawarehouse, types of schemas star schema and snowflake

		schema, SQL statements, DDL, DML queries, advanced subqueries.
2	NoSQL Databases	A Brief History of NoSQL, ACID vs. BASE, CAP Theorem, The BASE, Comparison of SQL and NoSQL, Advantages and Disadvantages, Categories of NoSQL Databases (key-value, columnar, document, graph database) History, MongoDB Design Philosophy, Non-Relational Approach, JSON-Based Document Store, Performance vs. Features, Running the Database Anywhere, SQL Comparison. The Data Model, JSON and BSON, The Identifier (_id), Basic Querying, Create and Insert, Explicitly Creating Collections, Inserting Documents Using Loop, Inserting by Explicitly Specifying _id, Update, Delete, Read, Using Indexes, Using Conditional Operators, Regular Expressions, MapReduce, aggregate (), MongoDB Document Data Model Approach
3	Distributed Systems and Big Data Processing	Concept of distributed systems, transparencies in Distributed system, concept of middleware, Introduction to HDFS, Core components of Hadoop, Hadoop web UI, differences between Hadoop 1,2 and 3 architectures. HDFS shell commands Introduction to MapReduce concept, Pig interface, queries in Pig Latin, data types of Pig Latin. Pig grunt shell Architecture of Hive, Configuration and starting hive shell, features of Hive, components of hive: user interface, meta store, HiveQL process engine, execution engine, query execution in Hive, data types in Hive, types of tables in Hive, create table, load data in Hive from HDFS.
4	Advanced Next-Gen Databases	Columnar database: HBASE What are columnar databases, Architecture of HBASE, HMaster, Region server, Zookeeper components, Advantages, and disadvantages of HBase, HBASE shell commands, DDL commands, DML commands, security commands, queries. Graph database: Importance of graph database, Graph model, learning about node and relationships, creating a node, creating relationship between nodes, analyzing similarity between nodes, case study using Neo4j

Text Books:

1. Big Data Analytics: Concepts, Techniques, Tools and Technologies, by G. Sudha, M. Thangaraj, S. Suguna (Author), PHI, 30 June 2022
2. Professional NoSQL, Shashank Tiwari, ISBN: 978-0-470-94224-6, Wiley ,September 2011

Reference Books:

1. SQL & NoSQL Databases: Models, Languages, Consistency Options and Architectures for Big Data Management, by Andreas Meier (Author), Michael Kaufmann (Author)
2. Hadoop: The Definitive Guide, Tom White, O'Reilly Media, June 2009

Teaching Plan:			
Unit No.	Unit Title	Teaching Methods	No. of Lectures
1	Big Data Foundations and Traditional Databases	Classroom Teaching and ICT	15
2	NoSQL Databases	Classroom Teaching and ICT	15
3	Distributed Systems and Big Data Processing	Classroom Teaching and ICT	15
4	Advanced Next-Gen Databases	Classroom Teaching and ICT	15

**Syllabus of Courses of M.Sc. Data Science
Programme at Semester IV
with Effect from the Academic Year 2026-2027**

Nomenclature of the Course	Next Generation Databases Practical
Class	M.Sc. Data Science
Semester	IV
Course Code	26_SDSDP002
No. of Credits	2
Nature	Practical
Type	Major Mandatory
Highlight revision specific to employability/ entrepreneurship/ skill development	The Next Generation Databases Practicals course is designed to boost employability, entrepreneurship, and hands-on database skills. Students gain practical experience in RDBMS queries (DDL, DML, TCL), cloud data warehouses (Snowflake, BigQuery), Hadoop, Hive, HBase, Pig, MongoDB, and Graph Databases (Neo4j). Key skills include ETL, data wrangling, aggregation, advanced querying, and data import/export, preparing learners for real-world database management, analytics, and data-driven solution development.
Course Outcomes:	
The learner will be able to :	
CO1: To understand the core concepts of programming before starting to write new programs. Students should be able to develop logic for Problem Solving.	
CO2: To manage data by relying on data structures such as strings, arrays, files, lists, and Dictionaries. Exemplify the various levels of decision making on a program and implement a mix of loops, functions, and control flow to extract information from a data structure.	
Note: The following set of practical's could be implemented in Python / R / Power BI / Tableau / PostgreSQL / MySQL / Snowflake / BigQuery or any	
Syllabus:	
Sr. No.	Aim of the Practical
1	Perform DDL,DML and TCL queries in any RDBMS
2	Access any Datawarehouse in cloud platform like Snowflake, Bigquery and perform ETL, data exploration including visualisation
3	Hadoop shell commands
4	Wordcount implementation using Pig
5	Data wrangling using Pig
6	Create managed and external table in HIVE, view in Hadoop Web UI

7	Create a namespace in HBASE with tables. Perform DML queries on the tables
8	Create a document and collection in MongoDB, performing advanced queries with conditional operators and aggregate functions
9	Perform aggregation using Map() and Reduce() function in MongoDB
10	Creating nodes and relationships in graph database using cypher. Perform exploration queries
11	Export and Import Data between RDBMS and CSV/JSON files
12	Write a program to demonstrate Data Indexing in MongoDB

Teaching Plan:

Practicals	Unit Title	Teaching Methods	No. of Lectures
All	All Practical's	Lab Session with ICT	60

Syllabus of Courses of M.Sc. Data Science
Programme at Semester IV
with Effect from the Academic Year 2026-2027

Nomenclature of the Course	Optimization Methods for Data Science	
Class	M.Sc. Data Science	
Semester	IV	
Course Code	26_SDSDT003	
No. of Credits	4	
Nature	Theory	
Type	Major Mandatory	
Highlight revision specific to employability/ entrepreneurship/ skill development	The revised syllabus emphasizes mathematical modeling and algorithmic problem-solving to strengthen employability in data-driven and optimization-based industries. It also nurtures entrepreneurial innovation through the application of nature-inspired optimization techniques to real-world challenges.	
Course Outcomes:		
The learner will be able to: CO 1: Apply mathematical foundations to formulate and analyze optimization problems. CO 2: Formulate and solve linear and integer programming models for real-world problems. CO 3: Solve unconstrained and constrained optimization problems using classical methods. CO4: Implement nature-inspired algorithms to tackle complex non-linear optimization problems.		
Syllabus:		
Unit No.	Unit Title	Sub Titles (Learning Points)
1	Mathematical Foundation	Mathematical Foundation :Linear algebra and matrices , Vector space and eigen analysis, Elementary multivariable calculus, Optimization and Optimality, General Formulation of Optimization Problems.Critical points-introduction, classification (saddle points, local minima, and maxima),First-order and second-order necessary and sufficient conditions (FONC, SONC, SOSC) for optimization.
2	Linear and Integer Programming	Introduction to linear programming model, Applications of LP, Simplex method, Duality, Karmarkar's method, Integer Linear Programming, Applications of IP, LP Relaxation, Branch and Bound, Mixed Integer Programming, Applications of MIP.

3	Unconstrained and Constrained Optimization	One-dimensional search methods, Gradient-based methods , Conjugate direction and quasi-Newton methods, Lagrange theorem, Barycentric Concept Modality, Multimodal Optimization, Cross-Validation, Generalization Error, Sampling Strategies.
4	Non-Linear Problem and Nature-Inspired Algorithms	Non-linear constrained optimization models, Secant Method, Newton's Method, KKT conditions, Projection methods- BFGS Method, Introduction to SI, Ant and Bee Algorithms, Particle Swarm Optimization, Firefly Algorithm, Cuckoo Search, Bat Algorithm, Flower Pollination Algorithm.

Text Books:

1. Optimization Techniques and Applications with Examples by Xin-She Yang Wiley 3rd 2018 2.
2. Optimization Techniques by A.K. Malik, S.K. Yadav, S.R. Yadav I.K. International Publishing House 1st 2012

Reference Books:

1. An introduction to Optimization by Edwin P K Chong, Stainslaw Zak, Wiley Interscience, 2013 2.
2. Nonlinear Programming by Dimitri Bertsekas, Athena Scientific, 1999

Teaching Plan:

Unit No.	Unit Title	Teaching Methods	No. of Lectures
1	Mathematical Foundation	Classroom Teaching and ICT	15
2	Linear and Integer Programming	Classroom Teaching and ICT	15
3	Unconstrained and Constrained Optimization	Classroom Teaching and ICT	15
4	Non-Linear Problem and Nature-Inspired Algorithms	Classroom Teaching and ICT	15

Syllabus of Courses of M.Sc. Data Science
Programme at Semester IV
with Effect from the Academic Year 2026-2027

Nomenclature of the Course	Optimization Methods for Data Science Practical
Class	M.Sc. Data Science
Semester	IV
Course Code	26_SDSDP004
No. of Credits	2
Nature	Practical
Type	Major Elective
Highlight revision specific to employability/ entrepreneurship/ skill development	The revised practical syllabus focuses on hands-on implementation of optimization algorithms using Python, enhancing technical and analytical employability skills. It promotes entrepreneurial problem-solving through practical applications of optimization in real-world data science projects.
Course Outcomes:	
The learner will be able to:	
CO1: Understand and formulate real-world problems as mathematical optimization problems involving single and multiple variables.	
CO2: Apply analytical techniques such as differentiation, integration, and linear programming (Simplex, Duality) to solve optimization problems.	
CO3: Implement and analyze various classical and modern optimization algorithms (Newton's, Secant, Lagrange, KKT, BFGS, Gradient Descent) using Python in Jupyter Notebook.	
CO4: Explore and apply nature-inspired optimization methods (Genetic Algorithm, Particle Swarm Optimization, Flower Pollination Algorithm) to solve complex data science problems.	
Note: Software: Working on - https://jupyter.org/try-jupyter/	
Syllabus:	
Sr. No.	Aim of the Practical
1	Perform Matrix Operations and Compute Differentiation of Vectors and Matrices using Python.
2	Implement Integration of Vectors and Matrices using Symbolic Computation.
3	Solve Linear Programming Problems using the Simplex Algorithm and Analyze Duality.
4	Implement Newton's Method for Solving Non-linear Equations and Optimization Problems.
5	Implement the Secant Method for Root Finding in Optimization Functions.
6	Implement the Lagrange Multiplier Method for Constrained Optimization Problems.
7	Implementation of KKT theorem in Conditions for Nonlinear Optimization.
8	Implementation of Broyden–Fletcher–Goldfarb–Shanno (BFGS) Quasi-Newton Method.
9	Develop and Execute the Particle Swarm Optimization (PSO) Algorithm for Function Minimization.
10	Implement the Flower Pollination Algorithm (FPA) for Global Optimization.
11	Compare Gradient Descent and Stochastic Gradient Descent (SGD) for Function Optimization.

12	Apply Genetic Algorithm (GA) for Solving Function Optimization Problems.		
Prescribed Text/s (If any):			
Teaching Plan:			
Practical's	Unit Title	Teaching Methods	No. of Lectures
All	10 Practical's	Lab Session with ICT	60

Syllabus of Courses of M.Sc. Data Science
Programme at Semester IV
with Effect from the Academic Year 2026-2027

Nomenclature of the Course	Blockchain Technologies for Data Science	
Class	M.Sc. Data Science	
Semester	IV	
Course Code	26_SDSET005	
No. of Credits	2	
Nature	Theory	
Type	Major Mandatory	
Highlight revision specific to employability/ entrepreneurship/ skill development	The Fundamentals and Platforms of Blockchain course enhances employability, entrepreneurship, and practical blockchain skills. Students gain hands-on experience with Ethereum, Hyperledger Fabric, IPFS, and smart contracts, building decentralized applications for data provenance, secure sharing, and blockchain-based solutions. The course fosters job-ready expertise, innovation, and practical development of secure, data-driven, and decentralized systems.	
Course Outcomes:		
The learner will be able to:		
CO 1: gain a thorough understanding of blockchain technology, including its architecture, key concepts, and different platforms.		
CO 2: apply blockchain technology to ensure data integrity, security, and provenance in data science projects.		
Syllabus:		
Unit No.	Unit Title	Sub Titles (Learning Points)
1	Fundamentals and Platforms of Blockchain	Introduction to Blockchain Technology, Blockchain Architecture, Consensus Mechanisms, Smart Contracts, Blockchain Platforms Overview, Security and Privacy in Blockchain, Ethereum and Smart Contracts, Hyperledger Fabric, IPFS (InterPlanetary File System), BigchainDB, Chainlink and Oracles, Blockchain Development Tools, Blockchain Development Environment.

2	Applications of Blockchain in Data Science	Data Provenance and Auditing, Secure Data Sharing, Decentralized Machine Learning, Data Monetization with Blockchain, Future Trends and Challenges, Working with Hyperledger Fabric, Decentralized Storage with IPFS, Integrating Blockchain with Machine Learning, Data Provenance Implementation, Building a Data Marketplace, Hyperledger Fabric Chaincode, Blockchain and IoT Integration.	
<p>Text Books:</p> <ol style="list-style-type: none"> 1. Intelligent Data Analytics, IoT, and Blockchain. (2023). United States: CRC Press. 2. Peng, S. (2021). Blockchain for Big Data: AI, IoT and Cloud Perspectives. United States: CRC Press. <p>Reference Books:</p> <ol style="list-style-type: none"> 1. Blockchain, Big Data and Machine Learning: Trends and Applications. (2020). United States: CRC Press 2. Machine Learning, Blockchain Technologies and Big Data Analytics for IoTs: Methods, Technologies and Applications. (2022). United Kingdom: Institution of Engineering and Technology 			
Teaching Plan:			
Unit No.	Unit Title	Teaching Methods	No. of Lectures
1	Fundamentals and Platforms of Blockchain	Classroom Teaching and ICT	15
2	Applications of Blockchain in Data Science	Classroom Teaching and ICT	15

Syllabus of Courses of M.Sc. Data Science
Programme at Semester IV
with Effect from the Academic Year 2026-2027

Nomenclature of the Course	Blockchain Technologies for Data Science Practical
Class	M.Sc. Data Science
Semester	IV
Course Code	26_SDSEP006
No. of Credits	2
Nature	Practical
Type	Major Elective
Highlight revision specific to employability/ entrepreneurship/ skill development	The Blockchain Practical's course equips students with hands-on skills in Ethereum, Hyperledger Fabric, IPFS, and Solidity smart contracts. Through real-world projects—voting systems, asset transfer, car auctions, IoT tracking, and data provenance—students gain practical experience in deploying decentralized applications, managing secure data, and integrating blockchain with IoT. These practical's develop job-ready expertise, foster innovation, and strengthen skills for blockchain entrepreneurship and data-driven solutions.
Course Outcomes:	
The learner will be able to:	
CO1: Set up blockchain development environments, write and deploy smart contracts, and manage Hyperledger Fabric networks.	
CO2: To deploy chaincode on Hyperledger Fabric, create secure blockchain-based voting systems, and integrate blockchain with IoT for secure data management.	
Note: Software: Node.js, Remix IDE, Truffle, Ganache, MetaMask, Solidity, Git Bash, Docker, Hyperledger Fabric binaries, IPFS, TensorFlow, PyTorch.	
Syllabus:	
Sr. No.	Aim of the Practical
1	Installation of Ethereum, Hyperledger Fabric, Ganache, and other tools.
2	Understanding Remix IDE: Environment, Compilation, and Deployment.
3	Implementation of Variable Types and Access Specifiers in Solidity.
4	Working with Structs and Mappings in Solidity Smart Contracts.
5	Implementation of a Blockchain-based Voting System using Remix IDE.
6	Implementation of Ownership and Asset Transfer Mechanism using Remix IDE.
7	Developing a Blockchain-based Car Auction System in Solidity.
8	Developing a Blockchain-enabled IoT Asset Tracking System in Solidity.
9	Configuring and running a Hyperledger Fabric network.
10	Storing and retrieving data using IPFS.
11	Implementing a blockchain-based data provenance system.
12	Combining blockchain with IoT for secure data management.

Prescribed Text/s (If any):			
Teaching Plan:			
Practical's	Unit Title	Teaching Methods	No. of Lectures
All	10 Practical's	Lab Session with ICT	60

Syllabus of Courses of M.Sc. Data Science
Programme at Semester IV
with Effect from the Academic Year 2025-2026

Nomenclature of the Course	Financial Risk Analytics and Management	
Class	M.Sc. Data Science	
Semester	IV	
Course Code	26_SDSET007	
No. of Credits	2	
Nature	Theory	
Type	Major Electives	
Highlight revision specific to employability/ entrepreneurship/ skill development	Mastery in risk identification, quantification, and mitigation equips learners for roles in finance, consulting, and data-driven decision-making. Understanding financial risks and market dynamics empowers entrepreneurs to make informed strategic and funding decisions.	
Course Outcomes:		
The learner will be able to :		
CO 1: Students should be able address the challenges associated with financial risk (market risk, credit risk and operation risk) through quantitative models and statistical methods.		
CO 2: To provide an overview of data analytics application under financial risk analytics context.		
Syllabus:		
Unit No.	Unit Title	Sub Titles (Learning Points)
1	Role of analytics in Risk Management and Financial Markets	<p>Introduction to Risk analytics, steps in Risk analytics, benefits of Risk analytics, Financial Risk analytics, Current states of financial risk analytics</p> <p>Introduction to Financial Risk management. Introduction, Risk Management and its benefits, Types of Risks, Financial Markets, Types of Financial Risks, Market Risk, Credit/counterparty risk, Operational Risk, Model Risk, Risk and Risk Factors, Financial Risk Management, Steps in Risk Management Process</p> <p>Introduction to Financial Markets. Data from FRED, Yahoo, and other sources. Empirical characteristics of economic and financial time series. Bootstrapping confidence intervals.</p> <p>Term Structure of Interest Rates. Bond pricing, forward and yield curves. Estimating Non-linear regression splines. Applications.</p> <p>Market Risk. Quantile (i.e., Value at Risk) and coherent (i.e., Expected Shortfall) risk measures.</p>

2	Credit Risk, Operational Risk and Measuring Volatility	<p>Credit Risk. Hazard rate models, Markov transition probabilities Risk measures, Laplace simulation with FFT. Operational Risk and Extreme Finance. Generate frequency and severity of operational loss distributions. Estimating operational risk distribution parameters. Simulating loss distributions.</p> <p>Measuring Volatility. Measuring volatility. GARCH estimation. GARCH simulation. Measuring Value at Risk (VaR) and Expected Shortfall (ES).</p> <p>Portfolio Optimization. Combining risk management with portfolio allocations. Optimizing allocations. Simulating the efficient frontier.</p> <p>Aggregating Enterprise Risks. Enterprise risk management analytics and application.</p>
---	---	---

Prescribed Text/s (If any):

1. Analytics in Finance and Risk Management by Nga Thi Hong Nguyen, Shivani Agarwal, Ewa Ziemia Released December 2023, Publisher(s): CRC Press, ISBN: 9781003808664
2. Risk Management & Financial Institutions, John C. Hull, 6th Edition, Wiley
3. Financial Risk Modelling and Portfolio Optimization with R, Bernhard Pfaff, 8 August 2016, Print ISBN: 9781119119661 | Online ISBN: 9781119119692
| DOI: 10.1002/9781119119692 © 2016, John Wiley & Sons, Ltd+amazon

Teaching Plan:

Unit No.	Unit Title	Teaching Methods	No. of Lectures
1	Role of analytics in Risk Management And Financial Markets	Classroom Teaching and ICT	15
2	Credit Risk, Operational Risk and Measuring Volatility.	Classroom Teaching and ICT	15

Syllabus of Courses of M.Sc. Data Science
Programme at Semester IV
with Effect from the Academic Year 2025-2026

Nomenclature of the Course	Financial Risk Analytics and Management Practical
Class	M.Sc. Data Science
Semester	IV
Course Code	26_SDSEP008
No. of Credits	2
Nature	Practical
Type	Major Electives
Highlight revision specific to employability/ entrepreneurship/ skill development	Proficiency in R for financial modeling, simulation, and interactive app development enhances job readiness in data-driven finance and fintech sectors. Developing full-stack analytical applications fosters entrepreneurial skills and equips learners to create finance tools for business or startup use.
Course Outcomes:	
The learner will be able to :	
CO1: Apply R programming concepts, data structures, statistical techniques, and financial models to perform computations and visualize data for finance-related problems.	
CO2: Model and estimate the term structure of interest rates using nonlinear least squares and compare alternative model specifications to evaluate financial scenarios and market risk.	
Note: The following practical's will be implemented using Python and R	
Syllabus:	
Sr. No.	Aim of the Practical
1	R for Finance: R computations, data structures, financial, probability, and statistics calculations, visualization. Documentation with R Markdown.
2	More R Warm-Ups. Functions, loops, control bootstrapping, simulation, and more visualization
3	Term Structure and Splines: <ul style="list-style-type: none"> ● Start with statistical definitions and financial models of bond prices. ● Move into scenario and explore the possibilities. ● Build a financially informed model of the term structure of forward empirical rates. ● Estimate the model with nonlinear least squares. ● Compare and contrast two competing model specifications.
4	Market Risk: <ol style="list-style-type: none"> 1. Measure risks using historical and parametric approaches 2. Interpret results relative to business decisions 3. Visualize market risk
5	Credit Risk: <ul style="list-style-type: none"> ● Use actual transaction and credit migration data to examine relationships among default and explanations of credit-worthiness.

	<ul style="list-style-type: none"> ● Simulate default probabilities using Markov chains. ● Explore hazard rates and the probability of transitioning from one credit state to another.
6	<p>Operational Risk:</p> <ul style="list-style-type: none"> ● Define frequency and Severity ● Calculate risk of loss ● Calculate potential loss ● Fire losses ● Estimating the extremes
7	<p>Measuring Volatility:</p> <ul style="list-style-type: none"> ● Use a fix for volatility clustering ● Fit AR-GARCH models ● Simulate volatility from the AR-GARCH model ● Measure the risks of various exposures
8	<p>Portfolio Analytics:</p> <ul style="list-style-type: none"> ● Portfolio Optimization. ● Combine risk management with portfolio allocations. ● Optimizing allocations.
9	Perform Risk analysis using Monte Carlo Simulations.
10	<p>Build an App</p> <p>The application will contain four architectural layers.</p> <ul style="list-style-type: none"> - Analytics <ol style="list-style-type: none"> 1. Libraries used in app processes 2. Function that wraps analytical script 3. Inputs from UI layer to server layer 4. Outputs from server layer to UI layer - User Interface (UI) <ol style="list-style-type: none"> 1. Slide bars for user to input range of parameters 2. Plots to display results 3. Text to report results - Server <ol style="list-style-type: none"> 1. Run analytics with inputs from the UI and from a simulation function 2. Generate outputs for UI - Application generator run application function with UI and Server inputs
11	To build and deploy a Shiny-based interactive financial risk analysis app integrating Monte Carlo simulations and user-defined parameters.
12	To compare two competing term structure models for bond pricing using nonlinear least squares estimation and evaluate model performance.
<p>Prescribed Text/s (If any):</p> <ol style="list-style-type: none"> 1 . Financial Risk Analytics: Measurement, Management and Examples in R Kindle Edition by R. K. Arora (Author), Perna Lal, Willey,20th April 2022 2 . Financial Engineering Analytics: A Practice Manual Using R, William G. Foote,2018-01 09 	

Teaching Plan:			
Practicals	Unit Title	Teaching Methods	No. of Lectures
All	All Practical's	Lab Session with ICT	30

*Syllabus of Courses of M.Sc. Data Science
Programme at Semester IV
with Effect from the Academic Year 2026-2027*

Nomenclature of the Course	Legal Analytics	
Class	M.Sc. Data Science	
Semester	IV	
Course Code	26_SDSET009	
No. of Credits	2	
Nature	Theory	
Type	Major Elective	
Highlight revision specific to employability/ entrepreneurship/ skill development	<p>Enhances employability by developing industry-relevant skills in legal data handling, NLP, and machine learning, preparing students for roles in legal analytics and tech-driven law practices.</p> <p>Fosters entrepreneurial potential by enabling students to build data-driven legal solutions, automation tools, and predictive systems that address real-world challenges in the legal domain.</p>	
Course Outcomes:		
The learner will be able to :		
CO1: Collect, preprocess, and analyze legal data using statistical and NLP techniques, gaining insights to support decision-making in legal practice.		
CO2: Apply machine learning, advanced NLP, and predictive analytics techniques to legal data.		
Syllabus:		
Unit No.	Unit Title	Subtitles (Learning Points)
1.	Legal Analytics and Data Science	<p>Overview of Legal Analytics: Definition and scope, Importance and applications, Key stakeholders like lawyers, judges, law firms, and legal departments.</p> <p>Legal Data Sources: Types of legal data, Public vs. private legal data, Data collection methods and sources.</p> <p>Data Privacy and Ethics: Ethical considerations, Privacy laws and regulations, Data anonymization and security.</p> <p>Data Cleaning and Preparation: Handling missing data, Data normalization and transformation, Feature engineering for legal datasets.</p> <p>Exploratory Data Analysis (EDA): Descriptive statistics for legal data, Data visualization techniques, Identifying patterns and trends in legal data.</p> <p>Text Mining and Natural Language Processing (NLP): NLP techniques for legal text: tokenization, stemming,</p>

		lemmatization, Sentiment analysis and named entity recognition in legal documents.
2.	Machine Learning for Legal Analytics	<p>Machine Learning for Legal Analytics: Overview of supervised learning algorithms, Applications in legal analytics, Overview of unsupervised learning algorithms, Applications in legal analytics.</p> <p>Advanced NLP Techniques: Topic modeling (e.g., LDA), Deep learning for NLP (e.g., transformers, BERT), Legal applications.</p> <p>Predictive Analytics in Law: Introduction, Building predictive models for legal case outcomes, Sentencing and recidivism prediction, Litigation risk assessment.</p> <p>Legal Research and Document Analysis: Automating legal research, Contract analysis and review, E-discovery and information retrieval.</p> <p>Legal Technology and Tools: Overview of legal tech and applications, legal analytics platforms, Hands-on analytics, Integration of analytics in law practice management systems.</p>

Prescribed Text/s (If any)

1. Legal Analytics: The Application of Artificial Intelligence and Machine Learning to the Law by Daniel Martin Katz, Michael J. Bommarito II, and Josh Blackman, Cambridge University Press, 2019.
2. Legal Data Mining, Machine Learning, and Visualization edited by Marcello Di Bello, Julian T. S. Hou, and Al Kagan, Springer, 2020.

Other Learning Resources recommended:

1. Data Analytics and Artificial Intelligence in Legal Practice by Edward J. Walters, Routledge, 2020.
2. Pattern Recognition and Machine Learning by Christopher M. Bishop, Springer, 2006.
3. Speech and Language Processing (3rd Edition, Draft) by Daniel Jurafsky and James H. Martin, Pearson, 2023 – for NLP techniques applicable to legal text analysis.

Teaching Plan:

Unit No.	Unit Title	Teaching Methods	No. of Lectures
1	Legal Analytics and Data Science	Classroom Teaching and ICT	15
2	Machine Learning for Legal Analytics	Classroom Teaching and ICT	15

Syllabus of Courses of M.Sc. Data Science
Programme at Semester IV
with Effect from the Academic Year 2026-2027

Nomenclature of the Course	Legal Analytics Practical
Class	M.Sc. Data Science
Semester	IV
Course Code	26_SDSEP010
No. of Credits	2
Nature	Practical
Type	Major Elective
Highlight revision specific to employability/ entrepreneurship/ skill development	This course strengthens employability by equipping students with in-demand skills in Python, NLP, machine learning, and legal data analytics, enabling them to analyze real legal datasets and generate meaningful insights for tech-driven legal roles. It also fosters entrepreneurial thinking by empowering learners to create AI-based solutions such as contract analyzers, predictive models, and legal automation tools.
Course Outcomes:	
The learner will be able to :	
CO1: Collect, preprocess, and structure legal datasets for analysis using Python and NLP techniques.	
CO2: Perform EDA and sentiment analysis to uncover patterns and insights in legal data.	
CO3: Build and evaluate ML models for case outcome prediction, clustering, and topic identification.	
CO4: Apply advanced NLP techniques for NER, contract analysis, and legal question-answering systems.	
The following practical's will be implemented using Python with relevant libraries (NumPy, pandas, scikit-learn, NLTK, spaCy) or R with relevant packages (tm, tidytext) or any other suitable software.	
Syllabus:	
Sr. No.	Aim of the Practical
1	Data Collection and Preparation: Use web scraping techniques to collect a dataset of court case summaries from a legal database or website
2	Conduct EDA on a dataset of legal cases to identify patterns and trends. · Load the dataset into a panda DataFrame · Calculate summary statistics (mean, median, mode) for numeric columns · Visualize the distribution of case outcomes and the frequency of different types of cases
3	Apply text preprocessing techniques to a collection of legal documents · Tokenize the text into words and sentences · Perform stemming and lemmatization · Remove stop words and special characters
4	Analyze the sentiment of legal opinions using NLP techniques. · Preprocess the legal opinions text data. · Apply sentiment analysis using TextBlob and VADER. · Compare the sentiment scores and interpret the results.

5	<p>Develop a machine learning model to predict the outcome of court cases.</p> <ul style="list-style-type: none"> · Preprocess the dataset, encoding categorical variables. · Split the data into training and testing sets. · Train a classification model (e.g., logistic regression, decision tree) and evaluate its performance.
6	<p>Apply clustering algorithms to group similar legal documents.</p> <ul style="list-style-type: none"> · Preprocess the text data. · Use TF-IDF to convert text to numerical features. · Apply k-means clustering and visualize the clusters
7	<p>Use topic modeling to identify themes in legal documents.</p> <ul style="list-style-type: none"> · Preprocess the text data. · Apply Latent Dirichlet Allocation (LDA) to identify topics. · Interpret and label the discovered topics.
8	<p>Use NER to identify and extract entities (e.g., names, dates, organizations) from legal documents.</p> <ul style="list-style-type: none"> · Load and preprocess the legal text data. · Apply spaCy's NER model to extract entities. · Visualize the extracted entities and analyze their occurrences.
9	<p>Extract and analyze key clauses from a set of contracts.</p> <ul style="list-style-type: none"> · Load contract documents in text or DOCX format. · Preprocess the text data. · Use regex and NLP techniques to extract specific clauses (e.g., termination, confidentiality).
10	<p>Create a QA system that can answer questions using a dataset of legal documents.</p> <ul style="list-style-type: none"> · Preprocess the corpus of legal documents. · Fine-tune a pre-trained BERT model on the legal QA task. · Evaluate the system's performance on a set of legal questions.
11	<p>To measure and visualize the similarity between legal documents to identify precedent cases or related legal texts.</p> <ul style="list-style-type: none"> ● Load and preprocess legal documents ● Preprocess text data ● Convert text to numerical vector ● Compute similarity score and visualize
12	<p>To analyze relationships between legal cases using citation patterns and uncover influential cases in a legal corpus.</p> <ul style="list-style-type: none"> ● Extract case names and citations from legal documents or databases. ● Create a citation network with cases as nodes and citations as edges. ● Compute centrality measures (degree, betweenness, closeness) to find influential cases. ● Visualize the network using graphs to see case connections and clusters. ● Analyze citation patterns to identify highly cited cases and legal precedent trends.

Prescribed Text/s (If any):

1. Legal Analytics: The Application of Artificial Intelligence and Machine Learning to the Law. Cambridge University Press. Katz, D. M., Bommarito II, M. J., & Blackman, J. (2019).
2. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media. Bird, S., Klein, E., & Loper, E. (2009).
3. LexNLP Natural Language Processing and Information Extraction for Legal and Regulatory Texts. Michaels, A., & Bommarito II, M. J. (2018).
4. LEGAL-BERT: The Muppets Straight Out of Law School. Chalkidis, I., Fergadiotis, M., Androutsopoulos, I., Aletras, N., & Malakasiotis, P. (2020).
5. Scikit-learn: Machine Learning in Python. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011).
6. An Introduction to Statistical Learning with Applications in R (2nd ed.). Springer. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021).
7. R in Action: Data Analysis and Graphics with R* (2nd ed.). Manning Publications. Kabacoff, R. I. (2015).

Teaching Plan:

Practical's	Unit Title	Teaching Methods	No. of Lectures
All	All Practical's	Lab Session with ICT	60

Syllabus of Courses of M.Sc. Data Science
Programme at Semester IV
with Effect from the Academic Year 2026-2027

Nomenclature of the Course	RESEARCH PROJECT
Course Code	26_SDSPE011
Class	M.Sc. Data Science
Semester	IV
No of Credits	6
Nature	Practical
Type	Research Project
Relevance with Employability/ Entrepreneurship/ Skill development	The Research Project outcomes for M.Sc. DS students, encompassing research methodologies, statistical analysis, econometric techniques, and financial management strategies, are highly relevant for both employability and entrepreneurship. Proficiency in these areas enhances analytical and problem-solving skills, making graduates valuable assets to employers in various fields. For entrepreneurs, these skills are crucial for informed decision making, market analysis, and effective business planning, contributing to the success and sustainability of their ventures in a dynamic business environment.

A student is expected to devote at least 3 to 4 months of effort to the Research Project Implementation on the project they submitted as a proposal in Semester III. Students should submit a detailed research project implementation report at the time of viva. Students are not permitted to change the project they submitted as a proposal in Semester III.

Course Outcomes:

1. Identify the research problem and formulate objectives
2. Choose appropriate methodology with proper tools and techniques Advanced Research hypothesis testing methods should be used (e.g T test, ANOVA, Chi-square, etc.)
3. Analyze and interpret the data collected from different sources. 4. Make decision or find out conclusions on the basis of data analysis 5. Recall and list key research paradigms and methodologies in the subject. 6. Explain the principles of statistical analysis and their application in subject of research.

Guidelines for Documentation of Research Project Implementation in Semester –IV

Certified Spiral Bound Copy with Certificate is required to submit at the time of Viva Examination. A student should submit a Research Project Implementation report with the following details:

- **Title:** Title of the Research Project.
- **Objective:** A detailed objective of the proposal is needed.
- **Introduction/Background:**
- **Related works/Literature Survey:** A detailed survey of the relevant works done by others in the domain. The student is expected to refer to at least 20 recent (last five years) research papers in addition to
Textbooks and web links in the relevant topic.
- **Methodology:** A proper and detailed procedure of how to solve the problem discussed. It shall contain the techniques, tools, software, and data to be used.
- **Implementation details:** A description of how the project has been implemented. ● **Experimental setup and results:** A detailed explanation of how experiments were conducted, what software was used, and the results obtained. Details like screenshots, tables, and graphs can come here. ● **Analysis of the results:** A description of what the results mean and how they have been arrived at. Different performing measures or
Statistical tools used etc may be part of this.
- **Conclusion:** A conclusion of the project performed in terms of its outcome ● **Future enhancement:** A small description of what enhancement can be done when more time and resources are available
- **Program code:** The program code may be given as an appendix. The project documentation needs to be signed by the teacher in charge and head of the Department.

Scheme of Examination: -

Total Credits	Total Hours	Marks	
		Internal	External
06	180	60	90

Internal Examination

A) Continuous Internal Evaluation:

Method		Marks		
Internal Viva 1		30		
Methodology	Implementation	Total		
15	15	30		
Internal Viva 2		30		
Experimental setup and results	Analysis of the results	Code	Document	Total
05	05	15	05	30

External Examination

B) External Evaluation:

Method				Marks			
External Viva				90			
Introduction	Objectives	Methodology	Code/Model	Results	Documents	Viva	Total
05	05	10	25	25	10	10	90

Format

1st page (Main Page)

Title of the problem of the Project

A Project Submitted

To

**R. P. Gogate College of Arts & Science and
R.V. Jogalekar College of Commerce (Autonomous), Ratnagiri**

of

Master in Commerce / Art/ Science

Under the Faculty of Commerce /Art/ Science

By

Name of Student

Under the Guidance of

Name of the Guiding Teacher

**R. P. Gogate college of Arts & Science and
R.V. Jogalekar College of Commerce (Autonomous), Ratnagiri**
Advocate N.V. Joshi Road,

Near Ratnagiri District Court, Ratnagiri

Month and Year

On separate page

Index

Chapter No	Title of the Chapter	Page No.
01		
02		
03		
04		
05		

On separate page

Declaration by learner

I the undersigned Miss/Mr. [Name of the learner] here by, declare that work embodied in this Research project work titled forms my own contribution to the research work carried out under the guidance of [Name of the guiding teacher] is a result of my own research work and has not been previously submitted to any other University or College for any other Degree/ Diploma.

Wherever reference has been made to previous works of others, it has been clearly indicated as such and included in the bibliography.

I, here by further declare that all information of this document has been obtained and presented in accordance with academic rules and ethical conduct.

Name and Signature of the learner

Certified by

Name and signature of the Guiding Teacher

On separate page

Acknowledgment

(Model structure of the acknowledgement)

To list who all have helped me is difficult because they are so numerous and the depth is so enormous.

I would like to acknowledge the following as being idealistic channels and fresh dimensions in the completion of this Research project.

I thank the **R. P. Gogate college of Arts & Science and R.V. Jogalekar College of Commerce (Autonomous), Ratnagiri** for giving me opportunity to do this project.

I would like to thank my **Principal, Prof. Dr M. R. Sakhalkar Sir**, for providing the necessary facilities required for completion of this project.

I take this opportunity to thank our Coordinator, for his moral support and guidance.

I would also like to express my sincere gratitude towards my project guide whose guidance and care made the project successful.

I would like to thank my College Library, for having provided various reference books and magazines related to my project.

Lastly, I would like to thank each and every person who directly or indirectly helped me in the completion of the project especially my Parents and Peers who supported me throughout my project.

Date:

Place: Ratnagiri

(Head and BoS Chairperson)

Evaluation Scheme

For Theory Course:

The allocation of marks for the Internal Assessment and Semester End Examinations are as follows-

Evaluation for the course will be of 100 marks conducted in a 60: 40 pattern. 60 marks will be for semester end examination and 40 marks will be for internal evaluation.

A) Internal Assessment for 4 Credit Course: 40 % (40 Marks)

	Particulars	
	Mid-Term Class Test It should be conducted using any learning management system such as Moodle (Modular object-oriented dynamic learning environment) Question Paper Pattern for Periodical Class Test/ Online Examination Duration: 40 Minutes Match the Column / Fill in the Blanks / Multiple Choice Questions/ True or False / Answer in One or Two Lines (Concept based Questions) (1 Marks each)	
	Assignment/ Case study/ Presentations	
	Attendance	
	Behaviour	

B) Semester End Examination for 4 Credit Course: 60% (60 Marks)

Question Paper Pattern

Maximum Marks: 60

Questions to be set: 05

Duration: 2 Hours

Question no	Unit	Options	Marks
Q1	Unit 1	Any 2 out of 4	12
Q2	Unit 2	Any 2 out of 4	12
Q3	Unit 3	Any 2 out of 4	12
Q4	Unit 4	Any 2 out of 4	12
Q5	All Units	MCQ and Fill in the Blanks	12

A) Internal Assessment for 2 Credit Course: 40 % (20 Marks)

Sr. No.	Particulars	Marks
01	Mid-Term Class Test It should be conducted using any learning management system such as Moodle (Modular object-oriented dynamic learning environment) Question Paper Pattern for Periodical Class Test/ Online Examination Duration: 40 Minutes Match the Column / Fill in the Blanks / Multiple Choice Questions/ True or False / Answer in One or Two Lines (Concept based Questions) (1 Marks each)	10
02	Assignment/ Case study/ Presentations	05
03	Attendance & Behavior	05

B) Semester End Examination for 2 Credit Course: 60% (30 Marks)

Question Paper Pattern

Maximum Marks: 30

Questions to be set: 03

Duration: 1 Hour

Question No	Unit	Options	Marks
Q1	Unit 1	Any 2 out of 4	10
Q2	Unit 2	Any 2 out of 4	10
Q3	Unit 3	Any 2 out of 4	10

For Practical Course:

The allocation of marks for the Internal Assessment and Semester End Examinations are as follows-

Evaluation for the course will be of 100 marks conducted in a 60: 40 pattern. 60 marks will be for semester end examination and 40 marks will be for internal evaluation.

a. Internal Examination: 20 Marks

Sr. No	Criterion	Marks
1	Journals containing minimum 20 practical's which are timely completed with desired output	10
2	Attendance & Practical Performance	10

b. Semester End**Examination: 30 Marks****Duration: 2 hrs**

Sr. No	Criterion	Marks
1	One Practical Question OR Combination of Practical Questions OR Combination of Practical Question and Theory Question	25
2	Viva	05

Standard of Passing

The learner to pass a course shall have to obtain a minimum of 40% marks in aggregate for each course where the course consists of Internal Assessment & Semester End Examination. The learner shall obtain minimum of 40% marks (i.e. 16 out of 40) in the Internal Assessment and 40% marks in Semester End Examination (i.e. 24 out of 60) separately, to pass the course and minimum of Letter Grade “P” in the project component, wherever applicable to pass a particular semester. A learner will be said to have passed the course if the learner passes the Internal Assessment & Semester End Examination together.

Performance Grading:

Letter Grades and Grade Points

Semester GPA/ Program CGPA Semester/Program	% of Marks	Alpha-Sign / Letter Grade Result
9.00-10.00	90.0 -100	0 (Outstanding)
$8.00 \leq 9.00$	$80.0 \leq 90.0$	A+ (Excellent)
$7.00 \leq 8.00$	$70.0 \leq 80.0$	A (Very Good)
$6.00 \leq 7.00$	$60.0 \leq 70.0$	B+ (Good)
$5.50 \leq 6.00$	$55.0 \leq 60.0$	B (Above Average)
$5.00 \leq 5.50$	$50.0 \leq 55.0$	C (Average)
$4.00 \leq 5.00$	$40.0 \leq 50.0$	P (Pass)
Below 4.00	Below 40	F (Fail)
Ab (Absent)	-	Absent

Date:

Place:- Ratnagiri



The Chairperson

BoS of Data Science