

Deccan Education Society's
**Kirti M. Doongursee College of
Arts, Science and Commerce
(AUTONOMOUS)**



Affiliated to
UNIVERSITY OF MUMBAI

Syllabus for
Program: Masters of Science
Course: M.Sc Part-II
Subject: Data Science

Choice Based Credit System (CBCS)
with effect from
Academic Year 2025-2026

PROGRAM OUTCOMES

PO	Description
	A student completing Master's Degree in Science Program will be able to
PO1	Demonstrate an in-depth understanding of the relevant science discipline. Recall, explain, extrapolate, and organize conceptual scientific knowledge for execution and application and also to evaluate its relevance
PO2	Critically evaluate, analyze, and comprehend a scientific problem. Think creatively, experiment and generate a solution independently, check and validate it and modify if necessary.
PO3	Access, evaluate, understand, and compare digital information from various sources and apply it for scientific knowledge acquisition as well as scientific data analysis and presentation.
PO4	Articulate scientific ideas, put forth a hypothesis, design and execute testing tools and draw relevant inferences. Communicate the research work in appropriate scientific language.
PO5	Demonstrate initiative, competence, and tenacity at the workplace. Successfully plan and execute tasks independently as well as with team members. Effectively communicate and present complex information accurately and appropriately to different groups
PO6	Use an objective, unbiased and non-manipulative approach in the collection and interpretation of scientific data and avoid plagiarism and violation of Intellectual Property Rights. Appreciate and be sensitive to environmental and sustainability issues and understand its scientific significance and global relevance
PO7	Translate academic research into innovation and creatively design scientific solutions to problems. Exemplify project plans, use management skills, and lead a team in the planning and execution of a task
PO8	Understand the cross-disciplinary relevance of scientific developments and relearn and reskill to adapt to technological advancements.

PROGRAM SPECIFIC OUTCOMES

PSO	Description
	On completion of M.Sc. Data Science program, students will be able to:
PSO1	Identify, analyze and design solutions for data analytics problems using fundamental principles of mathematics, Statistics, computing sciences, and relevant domain disciplines
PSO2	Acquire the skills in handling data analytics programming tools towards problem solving and solution analysis for domain specific problems.
PSO3	Understand and commit to professional ethics, regulations, responsibilities, and norms of professional computing practices
PSO4	Understand the role of statistical approaches and apply the same to solve the real-life problems in the fields of data analytics.
PSO5	Ability to apply the advanced concepts of Big Data to create a platform to gain analytical skills which impact business decisions and strategies
PSO6	Apply critical thinking to evaluate and optimize the computational resources in data analytics
PSO7	Understand, and compare different data sources and apply it for scientific knowledge acquisition as well as scientific data analysis and presentation.
PSO8	Use research-based knowledge to provide sustainable solutions in different domain using multidisciplinary data analytical approach
PSO9	Communicate effectively on the basis of presenting their research work and gain knowledge on documentation and reports writing in a professional way
PSO10	Create a new design for the complex computational problems which meets the specific needs for environmental and societal impact domains
PSO11	Demonstrate the principles of data science and apply these in the multidisciplinary environments to manage project
PSO12	Gain hands-on experience with real datasets, work on team projects, and understand business-oriented data challenges in industry, academia, or government sectors

**Deccan Education Society's
Kirti M. Doongursee College
(Autonomous) Proposed
Curriculum as per NEP-2020
Year of implementation- 2025-2026
Name of the Department-Data Science**

Semester	Course Code	Course Title	Vertical	Credit
III	25DSMJ911	Next Generation Databases	Major	4
	25DSMJ91	Next Generation Databases – Practical	Major	2
	25DSMJ912	Deep Neural Network	Major	4
	25DSMJ92	Deep Neural Network- Practical	Major	2
	25DSMJ913	Data Compliance and Auditing	Major	2
	25DSEL91	Image and Video Analytics (2)	Elective	4
	25DSELP91	Image and Video Analytics (2)- Practical		
	25DSEL92	Econometrics and Finance (2)		
	25DSELP92	Econometrics and Finance (2)- Practical		
	25DSEL93	Data Science for Agriculture (2)		
25DSELP93	Data Science for Agriculture (2)-Practical			
25DSRP93	Research Proposal	RP	4	
	25DSMJT11	Data Engineering	Major	4
	25DSMJPT1	Data Engineering – Practical	Major	2
	25DSMJT12	Generative AI	Major	4
	25DSMJPT2	Generative AI – Practical	Major	2

IV	25DSELT21	Financial Risk Analytics and Management (2)	Elective	4
	25DSELP1	Financial Risk Analytics and Management (2)-Practical [OR]		
	25DSELT22	Blockchain Technologies for Data Science (2)		
	25DSELP2	Blockchain Technologies for Data Science (2)-Practical [OR]		
	25DSELT23 25DSELP3	Legal Analytics (2) Legal Analytics (2)-Practical		
	25DSOJTT3	Research Project	Field Project	6

SEMESTER III

Course Code	MAJOR SEM – III	Credits	Lectures/Week
25DSMJ911	PAPER I- Next Generation Databases	4	4
<p>Course Outcomes:</p> <p>After successful completion of this course, students would be able to</p> <p>CO1: Learn the fundamentals of Big Data, its applications in various domains and the technologies in Bigdata ecosystem.</p> <p>CO2: Know RDBMS model, importance of ACID properties, HDFS File Structure, Map Reduce Framework</p> <p>CO3: Execute the different types of NoSQL databases like document database, columnar database, and graph database. perform SQL queries</p> <p>CO4: Compare the different Database system and data warehouse, the architecture and working of Enterprise Data warehouse, ACID and BASE</p>			

Unit	Details	Lectures
I	<p>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.</p> <p>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.</p>	15
II	<p>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)</p> <p>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</p>	15

III	<p>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.</p>	15
IV	<p>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.</p> <p>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</p>	15
<p>References:</p> <ul style="list-style-type: none"> ● Big Data Analytics: Concepts, Techniques, Tools and Technologies, by G. Sudha, M. Thangaraj, S. Suguna (Author), PHI, 30 June 2022 ● Professional NoSQL, Shashank Tiwari, ISBN: 978-0-470-94224-6, Wiley, September 2011 ● Machine Learning: The Art and Science of Algorithms that Make Sense of Data Peter Flach Cambridge University Press 2012 ● SQL & NoSQL Databases: Models, Languages, Consistency Options and Architectures for Big Data Management, by Andreas Meier (Author), Michael Kaufmann (Author) ● Hadoop: The Definitive Guide, Tom White, O'Reilly Media, June 2009 		

Course Code	MAJOR SEM III	Credits	Lectures/ Week
25DSMJP91	Next Generation Databases-Practical	2	4
<p>Course Outcomes: After successful completion of this course, students would be able to CO1: Gain knowledge about bigdata ecosystem and various bigdata technologies CO2: Understand the different type of NoSQL systems and their implementations. CO3: Perform DDL, DML and TCL SQL queries on any RDBMS CO4: Compare different database systems and its implementations CO4: Analyze ETL operations on any public cloud platform</p>			
<p>Note: - MySQL/ PostgreSQL, Snowflake , Hadoop, MongoDB, Neo4j</p>			
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 visualization		
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		

Course Code	MAJOR SEM – III	Credits	Lectur es/We ek
25DSMJ91 2	PAPER II- Deep Neural Network	4	4

Course Outcomes:

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

CO1: Learn the architectures and workings of various deep learning models

CO2: Adapt with cutting-edge research in deep learning to formulate and solve novel research questions.

CO3: Implement deep learning solutions for complex problems in computer vision, natural language processing.

CO4: Analyze and optimize deep learning models using modern techniques and frameworks.

Unit	Details	Lecture s
I	Core Deep Learning Architectures Deep Neural Networks (DNNs) and Convolutional Neural Networks (CNNs) a) Fundamentals of DNNs: Architecture, training, Optimization Algorithms (Gradient Descent, Adam, RMSprop), Overfitting, Underfitting, and Regularization Techniques, Evaluation Metrics and challenges b) Detailed study of CNNs: Architectural elements, pooling, and convolutional operations c) Advanced CNN architectures: AlexNet, VGG, GoogLeNet, ResNet	15
II	Recurrent Neural Networks (RNNs) and Extensions a) Basics of RNNs: Architecture, backpropagation through time, challenges b) Advanced RNN structures: LSTM (Long Short-Term Memory), GRU (Gated Recurrent Units) c) Applications of RNNs in sequence analysis and prediction	15
III	Modern Deep Learning Techniques Autoencoders and Attention Mechanisms a) Autoencoders: Architecture, types (sparse, denoising, variational), applications b) Attention mechanisms: Basics, types, and significance in model performance c) Case studies on attention in neural networks	15
IV	Transformers and BERT a) Introduction to Transformers: Architecture, self-attention, positional encodings b) Detailed exploration of BERT (Bidirectional Encoder Representations from Transformers): Architecture, training, fine-tuning, applications c) Generative Adversarial Networks (GANs): Introduction, training algorithms, Conditional GANs, applications d) Deep Reinforcement Learning: Introduction to Reinforcement Learning, Deep Q Network, applications	15

References:

- Deep Learning by Ian Goodfellow, Yoshua Bengio, and Aaron Courville, MIT Press, 2016
- Neural Networks and Deep Learning: A Textbook by Charu C. Aggarwal, Springer, 2018
- Introduction to Linear Regression Analysis Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining Wiley India Pvt. Ltd 3rd 2016
- Pattern Recognition and Machine Learning by Christopher M. Bishop, Springer, 2006
- Machine Learning: A Probabilistic Perspective by Kevin P. Murphy, MIT Press, 2012
- Deep Learning for Computer Vision by Rajalingappaa Shanmugamani, Packt Publishing, 2018
- Transformers for Natural Language Processing by Denis Rothman, O'Reilly Media, 2021

Course Code	MAJOR SEM III	Credits	Lectures/Week
25DSMJ92	Deep Neural Network - Practical	2	4
Course Outcomes After successful completion of this course, students would be able to CO1: Recall the concepts of deep neural network models. CO2: Describe various deep learning architectures such as CNNs, RNNs, LSTMs, Autoencoders CO3: Apply deep learning techniques to problems in image processing, natural language processing, and sequential data analysis. CO4: Evaluate and optimize deep learning models using modern techniques such as dropout, batch normalization, and advanced optimizers.			
Note: - Python using libraries such as TensorFlow, Keras, and PyTorch can be used			
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 Autoencoders in TensorFlow
8	Implement Variational Autoencoder (VAE) to generate new images
9	Apply Transformer Model for Machine translation task.
10	Apply BERT for Text Classification

Course Code	MAJOR SEM – III	Credits	Lectures/Week
25DSMJ913	PAPER III- Data Compliance	2	4

Course Outcomes

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

CO1: Gain a comprehensive understanding of major data protection and privacy regulations such as GDPR, and their implications for businesses.

CO2: Know the data compliance and data auditing Information Security Performance Metrics.

CO3: Develop the ability to plan, conduct, and document data audits using industry-standard methodologies and tools to ensure data accuracy and compliance

CO4: Compare different Data Auditing Information Security Performance Metrics

Unit	Details	Lectures
I	<p>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,</p> <p>Eprivacy- Regulating the electronic communications sector, The relationship between data protection and e-Privacy,</p> <p>Operational Data Protection - The three layers of an organization, implementing data protection in the people layer, implementing data protection in the paper layer.</p> <p>Core Law of Data Protection- The principles of data protection, Lawfulness, fairness and transparency – the first principle, Purpose limitation – the second principle, Data minimization – the third principle.</p> <p>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.</p>	15
II	<p>Data Auditing 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</p>	15

	<p>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 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>	
--	---	--

References:

- 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
- Data auditing The Ultimate Step-By-Step Guide by Gerardus Blokdyk, 2018, 5STARCOOKS
- Assessing Information Security (strategies, tactics, logic and framework) by A Vladimirov, K.Gavrilenko, and A.Michajlowski, Addison-Wesley Professional , 2010 Professional Data Engineer DAN SULLIVAN Sybex - Wiley 3rd 2020

Course Code	ELECTIVE SEM – III	Credits	Lectures/Week
25DSEL91	Image and Video Analytics	2	4

Course Outcomes:

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

CO1: Recall the basic techniques used in image and video analysis, including filtering, segmentation, feature extraction, and classification.

CO2: Identify the role of neural networks, particularly CNNs, in automating complex image and video tasks like classification and detection.

CO3: Apply image processing techniques to preprocess images and videos for further analysis (e.g., noise reduction, edge detection, resizing).

CO4: Analyze the computational requirements of various algorithms in image and video processing and optimize for performance, especially in real-time applications.

Unit	Details	Lectures
------	---------	----------

I	Image Analysis Introduction to Image Processing: Overview of digital images and basic image processing concepts, Image acquisition and representation, Introduction to image processing tools (e.g., OpenCV, PIL) Image Enhancement Techniques: Histogram equalization, Filtering: smoothing, sharpening, Edge detection and image segmentation Feature Extraction and Description: Keypoint detection (SIFT, SURF), Feature descriptors (HOG, LBP), Object recognition and classification Advanced Image Analysis: Image transforms (Fourier, Wavelet), Morphological operations, Image stitching and panoramic images	15
II	Video Analytics Introduction to Video Processing: Video acquisition and representation, Basic concepts of video processing, Introduction to video analytics frameworks Motion Analysis: Optical flow, Motion detection and tracking, Background subtraction techniques Object Detection and Tracking: Algorithms for object detection (YOLO, SSD), Multi object tracking (SORT, Deep SORT), Real-time tracking systems Video Event and Activity Recognition: Temporal event detection, Activity recognition and anomaly detection, Video summarization techniques	15

References:

- Intelligent Image and Video Analytics, El-Sayed M. El-Alfy, George Bebis, Mengchu Zhou, Routledge, Taylor and Francis Group, CRC, 2023
- Machine Learning for Audio, Image and Video Analysis: Theory and Applications, Francesco Camastra, Alessandro Vinciarelli, Springer, 2015
- Computer Vision Using Deep Learning Neural Network Architectures with Python and Keras, Apress 2021 (UNIT-III, IV and V), Vaibhav Verdhhan, 2021

Additional References:

- “Image Processing, Analysis, and Machine Vision”, Milan Sonka, Vaclav Hlavac, Roger Boyle, 4th edition, Thomson Learning, 2013.
- Richard Szeliski, “Computer Vision: Algorithms and Applications”, Springer Verlag London Limited, 2011.
- “Video Analytics for Business Intelligence”, Caifeng Shan, Fatih Porikli, Tao Xiang, Shaogang Gong, Springer, 2012.
- “Computer Vision: A Modern Approach”, D. A. Forsyth, J. Ponce, Pearson Education, 2003.
- “Computer & Machine Vision”, E. R. Davies, (2012), Fourth Edition, Academic Press.

Course Code	ELECTIVE SEM III	Credits	Lectures/Week
25DSMJP92	Image and Video Analytics-Practical	2	4
Course Outcomes			
<p>After successful completion of this course, students would be able to</p> <p>CO1: Capture, manipulate, and format images and video frames using OpenCV</p> <p>CO2: Utilize algorithms to extract and match features between images.</p> <p>CO3: Implement and evaluate image enhancement and segmentation methods, Also face detection and event recognition for video summarization.</p> <p>CO4: Analyze complex image using transformations and morphological operations and motion detection and object tracking in dynamic video scenes.</p>			
Note: Access to Python and relevant libraries (OpenCV, scikit-image, etc.).			
1	Image Acquisition and Basic Processing: Acquire and manipulate images (resize, crop, format conversion) using OpenCV.		
2	Histogram Equalization: Enhance image contrast by applying histogram equalization techniques.		
3	Image Filtering and Edge Detection: Apply filters and detect edges using Sobel and Canny methods.		
4	Image Segmentation: Segment images using global and adaptive thresholding and k-means clustering.		
5	Feature Extraction and Matching: Extract keypoints and match features between images using SIFT or ORB.		
6	Image Transformation and Morphology: Perform Fourier/Wavelet transforms and apply morphological operations.		
7	Image Stitching and Panorama Creation: Stitch multiple images together to create a panoramic image.		
8	Video Capture and Frame Extraction: Capture video and extract frames at regular intervals using OpenCV.		
9	Motion Detection and Optical Flow: Detect motion and analyze optical flow using background subtraction and optical flow algorithms.		
10	Object Detection in Videos: Detect objects in video sequences using YOLO or SSD models.		

Course Code	ELECTIVE SEM – III	Credits	Lectures/Week
25DSEL92	Econometrics and Finance	2	2

Course Outcomes:

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

CO1: Learners will understand the concepts of Econometrics and finance, including terms like regression analysis, time series analysis, capital asset pricing model (CAPM), and risk-return trade-offs.

CO2: Understand the role of econometrics in analyzing economic and financial data and its application in making informed financial decisions.

CO3: Apply econometric techniques to analyze financial data, such as estimating relationships between variables using regression models.

CO4: Analyze financial data using econometric methods to identify trends, relationships, and anomalies in financial markets or economic variables.

Unit	Details	Lectures
I	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.	15
II	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.	15

References:

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

Course Code	ELECTIVE SEM III	Credits	Lectures/Week
25DSELP92	Econometrics and Finance - Practical	2	4
Course Outcomes			
After successful completion of this course, students would be able to			
CO1: Recall the concepts of Econometrics and Finance.			
CO2: Know the application of econometrics in Finance			
CO3: Use of Econometrics in financial modelling			
CO4: Analyze results of financial modeling			
Note: Working on - Micro soft Excel / R Programming / Python			
1	Uni-variate Regression Model		
2	Multivariate Regression Model		
3	Methods to detect heteroscedasticity and remedial measures.		
4	Methods to detect multicollinearity and remedial measures		
5	Methods to detect auto-correlation and remedial measures.		
6	Interpreting on Dummy variable		
7	Financial modelling: best practices in spreadsheet design.		
8	Model Design and structure		
9	Auditing and Testing		
10	Macros and Security		

Course Code	ELECTIVE SEM – III	Credits	Lectures/Week
25DSEL93	Data Science for Agriculture	2	2
<p>Course Outcomes</p> <p>After successful completion of this course, students would be able to</p> <p>CO1: Gain the knowledge in data collection, preprocessing, and analysis for agricultural data.</p> <p>CO2: Know the sustainability and environmental impacts in agriculture.</p> <p>CO3: Build predictive models and perform EDA specific to agriculture.</p> <p>CO4: Attribute data-driven solutions for agricultural challenges</p>			

Unit	Details	Lectures
I	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, Case studies of big data applications 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.	15
II	Machine Learning Fundamentals: Introduction to supervised and unsupervised learning, Common algorithms (regression, classification, clustering), Model evaluation and validation. Predictive Analytics in Agriculture: Yield prediction models, Disease and pest prediction, Weather forecasting and its impact on agriculture. 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. Case studies and real-world examples.	15

References:

- Data Science for Agriculture by Simon Stelling, April 14, 2020
- Data Science for Agriculture by David L. Bunge
- "Practical Data Science for Agriculture" by Kellee Koenig

Additional References:

- Data Science for Agricultural Innovation and Productivity by Hamidah Ibrahim, A Veena, K P Asha Rani, ISBN-13978-9815196191, Publication date 13 February 2024

Course Code	ELECTIVE SEM III	Credits	Lectures/Week
25DSELP93	Data Science for Agriculture- Practical	2	4
<p>Course Outcomes: After successful completion of this course, students would be able to CO1: Learn Exploratory Data Analysis (EDA) and Data Collection Techniques CO2: Utilize Remote Sensing and GIS CO3: Develop Predictive Models and Precision Agriculture Techniques CO4: Forecast Market Prices and Analyze Supply Chains</p>			
<p>Note: - Python, R, data analysis libraries (Pandas, NumPy, Scikit-learn), and data visualization tools (Matplotlib, Seaborn, Tableau).</p>			
1	Web Scraping for Agricultural Data: Scrape weather data, soil data, and crop yield data from relevant websites - Tools: Python, BeautifulSoup, Scrapy. Cleaning and Preprocessing Data: Handle missing values, outliers, and normalize data for further analysis. Exploratory Analysis of Crop Yield Data: Perform descriptive statistics and visualize data distributions, trends, and correlations.		
2	Weather Prediction Using Time Series Analysis: Predict future weather patterns using historical weather data.		
3	Crop Yield Prediction Using Machine Learning: Build a machine learning model to predict crop yields based on various factors (soil type, weather, etc.). Tools: Scikit-learn, XGBoost		
4	Analyzing Satellite Images for Crop Health Monitoring: Use satellite imagery to assess crop health and identify areas needing attention. Tools: QGIS, Python, OpenCV.		
5	Land Use Classification Using Remote Sensing Data: Classify land use patterns using remote sensing data and machine learning algorithms. Tools: Python, Scikit-learn, GDAL.		
6	Soil Moisture Prediction Using Sensor Data: Predict soil moisture levels using data from soil sensors and environmental factors.		
7	Optimizing Irrigation Systems Using Data Analytics: Analyze irrigation data to optimize water usage and improve crop yield.		
8	Plant Disease Detection Using Image Processing: Detect plant diseases using image processing techniques on leaf images.		
9	Creating Yield Maps Using GPS Data: Generate yield maps to visualize spatial variability in crop production.		
10	Price Forecasting for Agricultural Products: Predict market prices for agricultural products using historical price data and other relevant factors.		

Course Code	SEM – III	Credits	Lectures/Week
25DSRP93	Research Proposal	4	4

Introduction

The NEP 2020 has emphasized on the inclusion of research and development in Higher Education Institutions. As colleges are an integral part of the knowledge department and creation NEP 2020 has introduced the research component to quite a substantial degree at postgraduate level. The multidisciplinary, transdisciplinary and translational research culture is expected to be introduced at postgraduate level. Such research projects undertaken will obviously enhance the research productivity, collaboration at national and international level in various industries, government as well as community based organizations and agencies.

Course Outcomes:

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

CO1: Define well formulated specific objectives that help to develop the overall research methodology

CO2: Identify the research methods and literature review

CO3: Apply pre-learnt concepts to design research problem with the help of literature survey and develop a thorough research proposal.

CO4: Analyze the research gap by performing comparative study.

Guidelines for Research Proposal

A student is expected to devote at least 2 to 3 months of effort to the Research Project Proposal.

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 30 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.
- Significance / Scope of the work
- Conclusion
- References

Certified Spiral Bound Copy duly signed by Internal Guide and Head of the Department is required to submit at the time of External Viva Examination

Evaluation Scheme for Sem-III Research Proposal

Internal Examination A) Continuous Internal Evaluation:

Method			Marks	
Internal Viva 1			25	
Topic Weightage	Introduction	Objectives	Literature Survey	Total
05	05	05	10	25
Internal Viva2			25	
Proposed Methodology	Significance / Scope and Conclusion	Documentation	Total	
10	05	10	25	

External Examination A) External Evaluation:

Method				Marks			
External Viva				50			
Topic Weightage	Introduction	Objectives	Literature Review	Proposed Methodology	Documentation	Presentation/ Viva	Total
05	05	05	10	05	10	10	50

SEMESTER IV

Course Code	MAJOR SEM – IV	Credits	Lectures/Week
25DSMJ911	PAPER I- Data Engineering	4	4

Course Outcomes:

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

CO1: Identify key components and services in modern data engineering ecosystems, including GCP tools such as BigQuery, Cloud SQL, Cloud Storage, Cloud Spanner, and others and Recall different types of data structures, storage models, schema design principles, and processing strategies.

CO2: Explain how business needs drive the selection of data storage and processing technologies, describe the architecture of cloud-based data pipelines, and outline the importance of security, privacy, and legal compliance in data engineering using tools like Cloud IAM and the DLP API.

CO3: Design and implement end-to-end data solutions by leveraging GCP services for storage and processing, applying best practices for resource management, and utilizing tools like Data Studio and Datalab for data visualization and exploration.

CO4: Evaluate schema design choices for scalability, availability, and reliability in GCP services, analyze and optimize the migration of Hadoop and Spark infrastructures, assess data security and compliance in workflows, and compare machine learning pipeline structures and deployment options for data-driven solutions.

Unit	Details	Lectures
I	<p>Selecting Appropriate Storage Technologies: From Business Requirements to Storage Systems, Technical Aspects of Data, Types Of Structure, Schema Design Consideration</p> <p>Building and Operationalizing Storage Systems: Cloud SQL, Cloud Spanner, Cloud Bigtable, Cloud Firestore, BigQuery, Cloud Memorystore, Cloud Storage, Unmanaged Databases</p> <p>Designing Data Pipelines: Overview Of Data Pipelines, GCP Pipeline Components, Migrating Hadoop and Spark To GCP</p> <p>Designing a Data Processing Solution: Designing Infrastructure, Designing for Distributed Processing, Migrating a Data Warehouse</p>	15
II	<p>Building and Operationalizing Processing Infrastructure: Provisioning and Adjusting Processing Resources, Monitoring Processing Resources</p> <p>Designing for Security and Compliance: Identity and Access Management with Cloud IAM, Using IAM with Storage and Processing Services, Data Security, Ensuring Privacy with the Data Loss Prevention API, Legal Compliance</p>	15

III	Building and Operationalizing Processing Infrastructure: Provisioning and Adjusting Processing Resources, Monitoring Processing Resources, Designing for Security and Compliance: Identity and Access, Management with Cloud IAM, Using IAM with Storage and Processing Services, Data Security, Ensuring Privacy with the Data Loss Prevention API, Legal Compliance
IV	Designing Databases for Reliability, Scalability, and Availability: Designing Cloud Bigtable Databases for Scalability and Reliability, Designing Cloud Spanner Databases for Scalability and Reliability, Designing BigQuery Databases for Data Warehousing, Understanding Data Operations for Flexibility and Portability: Cataloging and Discovery with Data Dataprep, Visualizing with Data Studio, Exploring Data with Cloud Datalab, Orchestrating Workflows with Cloud Composer, Deploying Machine Learning Pipelines: Structure of ML Pipelines, GCP Options for Deploying Machine Learning PipelineCatalog, Data Preprocessing

References:

- Professional Data Engineer DAN SULLIVAN Sybex - Wiley 3rd 2020
- Data Driven Science and Engineering STEVEN L. BRUNTON, J. NATHAN KUTZ Cambridge University Press 2nd 2019
- Data Security in Cloud Computing Vimal Kumar, Sivadon Chaisiri and Ryan Ko The Institution of Engineering and Technology 2020
- Data Engineering on Azure Vlad Riscutia Manning Publications 2021

Course Code	MAJOR SEM IV	Credits	Lectures/ Week
25DSMJPT1	Data Engineering-Practical	2	4

Course Outcome:

After successful completion of this course, students would be able to
CO1: Recall the different types of databases and storage systems used in data engineering, such as relational databases, NoSQL, Hadoop, and cloud-based data platforms.

CO2: Understand the role of data engineering in the data lifecycle, and how it supports data analytics, machine learning, and business intelligence.

CO3: Apply data engineering principles to ensure data quality, perform data cleaning, and handle missing or inconsistent data effectively.

CO4: Analyze different designs of databases to ensure reliability and scalability

1	Perform the Given Operations in CouchDB.
2	Perform the Practical using MongoDB.
3	Create a AutoML Project on Car Price Prediction in AzureML.

4	Data Exploration Using PySpark SQL.
5	Perform Distributed Data Processing Using PySpark RDD.
6	Data Cleaning and Transformation Using ML Pipeline.
7	Perform Classification using ML Pipeline with PySpark in Databricks.
8	Perform Regression using ML Pipeline with PySpark in Databricks on US cars dataset.
9	Perform Clustering with PySpark.
10	Perform Deep Learning in Keras in Distributed computing using PySpark.

Course Code	MAJOR SEM – IV	Credits	Lectur es/We ek
25DSMJ912	PAPER II– Generative AI	4	4

Course Outcomes:

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

CO1: Listing the different types of generative models (e.g., GANs, VAEs, autoregressive models) and their applications.

CO2: Know GANs, VAEs, Diffusion and NLP based Generative Models.

CO3: Execute different techniques of generative model to create new content

CO4: Analyze the structure of a generative AI model, such as identifying the components of a GAN

Unit	Details	Lectures
I	<p>Introduction to Generative AI</p> <ul style="list-style-type: none"> • Basics of AI & Machine Learning • Introduction to Generative AI and its applications • Types of Generative Models: GANs, VAEs, Diffusion Models • Overview of NLP-based Generative Models (GPT, BERT, T5) • Ethical considerations in Generative AI 	15
II	<p>Generative Adversarial Networks (GANs)</p> <ul style="list-style-type: none"> • Architecture of GANs: Generator & Discriminator • Training process & challenges (Mode Collapse, Vanishing Gradients) • Variants of GANs: DCGAN, CycleGAN, StyleGAN • Applications of GANs: Image synthesis, Data Augmentation • Implementation of a basic GAN using TensorFlow/PyTorch 	15
III	<p>Variational Autoencoders (VAEs) and Diffusion Models</p> <ul style="list-style-type: none"> • Autoencoders vs. VAEs • Latent space representation & sampling • Introduction to Diffusion Models • Training and fine-tuning VAEs and Diffusion Models • Applications: Image Denoising, Text-to-Image Generation 	15

IV	NLP-Based Generative Models <ul style="list-style-type: none"> • Transformers and Attention Mechanism • Pretrained Language Models (GPT, BERT, LLaMA, etc.) • Fine-tuning Generative Models for specific tasks • Text Generation, Summarization, and Conversational AI • AI Safety and Bias in Generative AI 	15
References: <ul style="list-style-type: none"> • "Probabilistic Machine Learning: An Introduction" by Kevin P. Murphy • "Artificial Intelligence: A Modern Approach" by Stuart Russell and Peter Norvig 		

Course Code	MAJOR SEM IV	Credits	Lectures/Week
25DSMJ21	Generative AI -Practical	2	4
Course Outcomes: After successful completion of this course, students would be able to CO1: Recall the concepts of GANs,VAEs, Diffusion and NLE based generative models. CO2: Identify the architectures of different generative AI models and LLMs. CO3: Implement the different Generative AI model for new content creation. CO4: Compare different Generative AI Models.			

1	Prompt Engineering To get the best possible responses from an AI model- <ul style="list-style-type: none"> • Specificity & Clarity • Context Retention • Role-based Prompting • Constraint Testing • Chain-of-Thought Prompting • Creative vs. Analytical Responses
2	Identify different LLMs Getting introduced with different LLMs
3	Introduction to Generative Models Implement a simple autoencoder using PyTorch/TensorFlow.
4	Building a GAN from Scratch Train a basic GAN to generate handwritten digits (MNIST dataset).
5	Improving GAN Training Implement DCGAN and experiment with different architectures.
6	Text Generation with GPT Use OpenAI's GPT-3/4 API to generate text based on a given prompt.

7	Image Style Transfer using CNNs & GANs Implement Neural Style Transfer using VGG-19.
8	Training a Variational Autoencoder (VAE) Train a VAE on the CelebA dataset for face generation.
9	Diffusion Model for Image Generation Implement a basic diffusion model and generate high-quality images.
10	Fine-tuning a Pretrained Language Model Fine-tune GPT/BERT for text summarization or chatbots.

Course Code	Elective SEM IV	Credits	Lectures/Week
25DSELT21	Financial Risk Analytics and Management	2	4

Course Outcomes:

After successful completion of this course, students would be able to
CO1: Recall the knowledge of handling financial risks using data analytics framework.
CO2: Know the data analytics application under financial risk analytics context.
CO3: Implement appropriate risk measures in the financial domain.
CO4: Analyze the challenges associated with financial risk (market risk, credit risk and operational risk) through quantitative models and statistical methods.

Unit	Details	Lectures
I	<p>Role of analytics in Risk Management: 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. Market Risk. Quantile (i.e., Value at Risk) and coherent (i.e., Expected Shortfall) risk measures.</p>	15

II	<p>Credit Risk: Hazard rate models, Markov transition probabilities Risk measures, Laplace simulation with FFT.</p> <p>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>	15
<p>References:</p> <ul style="list-style-type: none"> Financial Engineering Analytics: A Practice Manual Using R, William G. Foote, 2018-01-09 Risk Management & Financial Institutions, John C. Hull, 6th Edition, Wiley Financial Risk Modelling and Portfolio Optimization with R, Bernhard Pfaff, 8 August 		

Course Code	Elective SEM IV	Credits	Lectures/ Week
25DSELP1	Financial Risk Analytics and Management - Practical	2	4

<p>Course Outcomes:</p> <p>After successful completion of this course, students would be able to</p> <p>CO1: Learn the concepts of financial risk Analytics</p> <p>CO2: Summarize data analytics tools in handling Financial risk domain.</p> <p>CO3: Implement data analytics techniques to on financial sector dataset.</p> <p>CO4: Analyze the existing and potential data analytic solutions in financial risk management.</p>			
---	--	--	--

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	<p>Term Structure and Splines:</p> <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	<p>Market Risk:</p> <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	<p>Credit Risk:</p> <ul style="list-style-type: none"> ● Use actual transaction and credit migration data to examine relationships among default and explanations of credit-worthiness. ● 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	<p>Perform Risk analysis using Monte Carlo Simulations.</p>
10	<p>Build an App</p> <p>The application will contain four architectural layers.</p> <ul style="list-style-type: none"> - Analytics 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

	<ul style="list-style-type: none"> - 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
--	---

Course Code	Elective SEM IV	Credits	Lectures/Week
24DSELT22	Blockchain Technologies for Data Science	2	4

Course Outcomes:

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

CO1: Gain a thorough understanding of blockchain technology, including its architecture, key concepts, and different platforms.

CO2: Describe IPFS for decentralized storage and integrating blockchain to ensure data integrity in machine learning models

CO3: Apply blockchain technology to ensure data integrity, security, and provenance in data science projects.

CO4: Provide innovative blockchain-based solutions for real world data science problems, including decentralized storage, data marketplaces, and secure data sharing.

Unit	Details	Lectures
I	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, Setting Up Blockchain Development Environment, Writing Smart Contracts in Solidity, Deploying Smart Contracts on Ethereum	15
II	Applications of Blockchain in Data Science: Data Provenance and Auditing, Secure Data Sharing, Decentralized Machine Learning, Data Monetization with Blockchain, Case Studies, Future Trends and Challenges, Working with Hyperledger Fabric, Decentralized Storage with IPFS, Integrating Blockchain with Machine Learning, Data Provenance Implementation,	15

	Building a Data Marketplace, Hyperledger Fabric Chaincode, Blockchain-based Voting System, Blockchain and IoT Integration	
<p>References:</p> <ul style="list-style-type: none"> Intelligent Data Analytics, IoT, and Blockchain. (2023). United States: CRC Press. Peng, S. (2021). Blockchain for Big Data: AI, IoT and Cloud Perspectives. United States: CRC Press. <p>Additional References:</p> <ul style="list-style-type: none"> Blockchain, Big Data and Machine Learning: Trends and Applications. (2020). United States: CRC Press. Machine Learning, Blockchain Technologies and Big Data Analytics for IoTs: Methods, Technologies and Applications. (2022). United Kingdom: Institution of Engineering and Technology 		

Course Code	Elective SEM IV	Credits	Lectures/Week
25DSELP22	Blockchain Technologies for Data Science -Practical	2	4

Course Outcomes:

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

CO1: Set up blockchain development environments, write and deploy smart contracts, and manage Hyperledger Fabric networks.

CO2: Acquire hands-on experience in developing and deploying smart contracts, setting up blockchain networks, and integrating blockchain with data science applications.

CO3: Implement blockchain-based data provenance systems and develop decentralized data marketplaces.

CO4: Organize secure data management by integrating blockchain with IoT.

Note: - Node.js, npm, Truffle, Ganache, MetaMask, Solidity, Git Bash, Docker, Hyperledger Fabric binaries, IPFS, TensorFlow, PyTorch,

1	Installation of Ethereum, Truffle, Ganache, and other tools
2	Writing and deploying basic smart contracts on Ethereum
3	Configuring and running a Hyperledger Fabric network

4	Storing and retrieving data using IPFS
5	Using blockchain to ensure data integrity in ML models
6	Implementing a blockchain-based data provenance system
7	Developing a decentralized marketplace for data exchange
8	Writing and deploying chaincode on Hyperledger Fabric
9	Implementing a secure voting system using blockchain
10	Combining blockchain with IoT for secure data management

Course Code	Elective SEM IV	Credits	Lectures/Week
25DSELT23	Legal Analytics	2	4

Course Outcomes:

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

CO1: Recall hands-on experience with legal analytics tools and platforms, integrating analytics into law practice management.

CO2: Explain the ethical considerations and legal regulations surrounding data privacy in the context of legal analytics.

CO3: Execute text mining and NLP techniques to legal texts to extract meaningful insights

CO4: Analyze advanced NLP techniques such as topic modeling and deep learning for complex legal text analysis

Unit	Details	Lectures
I	Overview of Legal Analytics: Definition and scope of legal analytics. Importance and applications in the legal field, Key stakeholders: lawyers, judges, law firms, and legal departments. Legal Data Sources: Types of legal data: case law, statutes, regulations, and legal documents, Public vs. private legal data, Data collection methods and sources (e.g., legal databases, web scraping)	15

	<p>Data Privacy and Ethics: Ethical considerations in legal analytics, Privacy laws and regulations (e.g., GDPR, CCPA), 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 (e.g., histograms, box plots, word clouds), Identifying patterns and trends in legal data</p> <p>Text Mining and Natural Language Processing (NLP): NLP techniques for legal text: tokenization, stemming, lemmatization, Sentiment analysis and named entity recognition in legal documents</p>	
<p>II</p>	<p>Machine Learning for Legal Analytics: Supervised Learning: Overview of supervised learning algorithms, Applications in legal analytics: case outcome prediction, legal document classification. Unsupervised Learning: Overview of unsupervised learning algorithms, Applications in legal analytics: clustering similar cases, anomaly detection Advanced NLP Techniques: Topic modeling (e.g., LDA), Deep learning for NLP (e.g., transformers, BERT), Legal question-answering systems and chatbots Predictive Analytics in Law: Building predictive models for legal case outcomes, Sentencing and recidivism prediction, Litigation risk assessment Legal Research and Document Analysis: Automating legal research, Contract analysis and review, E-discovery and information retrieval. Legal Technology and Tools: Overview of legal tech tools and platforms, Hands-on with legal analytics software (e.g., Lex Machina, Ravel Law), Integration of legal analytics in law practice management systems.</p>	<p>15</p>
<p>References:</p> <ul style="list-style-type: none"> • 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 • Legal Data Mining, Machine Learning, and Visualization edited by Marcello Di Bello, Julian T. S. Hou, and Al Kagan, Springer, 2020 		

Course Code	Elective SEM IV	Credits	Lectures/ Week
25DSELP23	Legal Analytics Practical	2	4
<p>Course Outcomes:</p> <p>After successful completion of this course, students would be able to</p> <p>CO1: Recall Legal Analytics Concepts.</p> <p>CO2: Use data visualization tools to represent legal data and insights effectively.</p> <p>CO3: Apply data analytics to support legal decision-making processes</p> <p>CO4: Analyze legal data using deep learning model</p>			

1	Data Collection and Preparation: Use web scraping techniques to collect a dataset of court case summaries from a legal database or website
2	<p>Conduct EDA on a dataset of legal cases to identify patterns and trends.</p> <ul style="list-style-type: none"> • 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	<p>Apply text preprocessing techniques to a collection of legal documents</p> <ul style="list-style-type: none"> • Tokenize the text into words and sentences • Perform stemming and lemmatization • Remove stop words and special characters
4	<p>Analyze the sentiment of legal opinions using NLP techniques.</p> <ul style="list-style-type: none"> • 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> <p>Preprocess the text data.</p> <ul style="list-style-type: none"> • 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	Use NER to identify and extract entities (e.g., names, dates, organizations) from

	<ul style="list-style-type: none"> • legal documents. • Load and preprocess the legal text data. • Apply spaCy's NER model to extract entities. <p>Visualize the extracted entities and analyze their occurrences.</p>
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

Course Code	Course title	Credits	Lectur es/We ek
25DSOJTT3	RESEARCH PROJECT	6	4

Introduction

The research project proposal needs to be implemented by following the given timeline Under the NEP, students are expected to get industry ready by the time they pass out of their Masters' degree course. There is also an emphasis on research so that every student is expected to carry out independent research project as a part of their Post Graduate program. The emphasis is on research that is socially applicable, and carried out with scientific rigor. One benchmark of good quality research is publication of the project either in International or National level scientific journals or the presentation of students' research work at International, National or State level conferences. With this broad objective, the following has been proposed for student research projects at Masters level.

Course Outcomes:

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

CO1: Facilitate substantial data collection for the proposed research work

CO2: Carry out research following ethical aspects of research activities.

CO3: Develop the proposed system

CO4: Analyze the research findings/conclusions / results obtained

Note: A student is expected to devote at least 3 to 4 months of effort to the Research Project Implementation on the proposal submitted 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.

Guidelines for Documentation of Research Project Implementation in Semester – IV

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 30 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.

Certified Black Bound Copy duly signed by Internal Guide and Head of the Department is required to submit at the time of External Viva Examination

Evaluation Scheme for Sem-IV Research Project

Internal Examination A) Continuous Internal Evaluation:

Method		Marks		
Internal Viva 1				
Methodology	Implementation	Total		
20	20	40		
Internal Viva 2				
Experimental setup and results	Analysis of the results	Code	Document	Total
05	10	15	05	35

External Examination A) External Evaluation:

Method				Marks			
External Viva				75			
Introduction	Objectives	Methodology	Code/Mode	Results	Document	Viva	Total
05	05	10	20	20	05	10	75

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

I. Internal Evaluation for Theory Courses – 40 Marks

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

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

II. External Examination for Theory Courses – 60 Marks

Duration: 2 Hours

Theory question paper pattern:

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

- All questions shall be compulsory with internal choice within the questions.
- Each Question may be subdivided into sub questions as a, b, c, d, etc. & the allocation of Marks depends on the weightage of the topic.
- Paper Pattern of Theory Paper:(for 4 credits)

DES's Kirti M. Doongursee College (AUTONOMOUS), Dadar (W), Mumbai-28		
Regular / Additional / ATKT Examination		
Duration: 2 Hour		Max Marks: 60
Date: <i>(For office use)</i>	Time:	Code:
N. B.	i)	<i>All Questions are compulsory.</i>
	ii)	<i>Mixing of sub-questions is not allowed</i>
	iii)	<i>Draw neat labeled diagrams wherever necessary.</i>
Q. No.		Marks
Q.1 A		07
OR		
Q.1 B		07
Q.1 C		08
OR		
Q.1 D		08
Q.1 E		07
OR		

Q.1 F		07
Q.2 G		08
OR		
Q.2 H		08
Q.3 I		07
OR		
Q.3 J		07
Q.3 K		08
OR		
Q. 3 L		08
Q.4 M		07
OR		
Q.4 N		07
Q.4 O		08
OR		
Q.4 P		08

III. Practical Examination

- Each core subject carries 50 Marks
- Duration: 2 Hours for each practical course.
- Minimum 80% practical from each core subjects are required to be completed.
- Certified Journal is compulsory for appearing at the time of Practical Exam

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

Evaluation Scheme for Second Year (PG) under NEP (2 credits)

I. Internal Evaluation for Theory Courses – 20 Marks

- 1) Continuous Internal Assessment(CIA) Assignment - Tutorial/ Case Study/ Project / Presentations/ Group Discussion / Ind. Visit. – 10 marks
- 2) Continuous Internal Assessment(CIA) ONLINE Unit Test – 10 marks

II. External Examination for Theory Courses – 30 Marks

Duration: 1 Hours

Theory question paper pattern:

Question	Based on	Marks
Q.1	Unit I	15
Q.2	Unit II	15

- All questions shall be compulsory with internal choice within the questions.
- Each Question may be subdivided into sub questions as a, b, c, d, etc. & the allocation of Marks depends on the weightage of the topic.
- Paper Pattern of Theory Paper:(for 2 credits)

DES's Kirti M. Doongursee College (AUTONOMOUS), Dadar (W), Mumbai-28		
Regular / Additional / ATKT Examination		
Duration: 1 Hour		Max Marks: 30
Date:	Time:	Code:
<i>(For office use)</i>		
N. B.	i)	<i>All Questions are compulsory.</i>
	ii)	<i>Mixing of sub-questions is not allowed</i>
	iii)	<i>Draw neat labeled diagrams wherever necessary.</i>
Q. No.		Marks
Q.1 A		07
OR		
Q.1 B		07
Q.1 C		08
OR		
Q.1 D		08
Q.1 E		07
OR		

Q.1 F		07
Q.2 G		08
OR		
Q.2 H		08

III. Practical Examination

- Each core subject carries 50 Marks.
- Duration: 2 Hours for each practical course.
- Minimum 80% practical from each core subjects are required to be completed.
- Certified Journal is compulsory for appearing at the time of Practical Exam

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