University of Mumbai

No. AAMS_UGS/ICC/2024-25/ 100

CIRCULAR:-

Attention of the Principals of the Affiliated Colleges, Directors of the Recognized Institutions and the Head, University Departments is invited to this office circular No. AAMS_UGS/ICC/2023-24/23 dated 08th September, 2023 relating to the NEP UG & PG Syllabus.

They are hereby informed that the recommendations made by the **Ad-hoc Board of Studies in Data Science** at its meeting held on 02nd July, 2024 and subsequently passed by the Board of Deans at its meeting held on 10th July, 2024 **vide** item No. 6.1 (N) have been accepted by the Academic Council at its meeting held on 12th July, 2024 **vide** item No.6.1 (N) and that in accordance therewith **syllabus** for the **M.Sc (Data Science) (Sem. III & IV)** is introduced as per appendix (NEP 2020) with effect from the academic year 2024-25.

(The circular is available on the University's website www.mu.ac.in).

Baliram

MUMBAI – 400 032
22nd August, 2024

(Prof.(Dr) Baliram Gaikwad)
I/c Registrar

To

The Principals of the Affiliated Colleges, Directors of the Recognized Institutions and the Head, University Department.

A.C/6.1(N)/12/07/2024

Copy forwarded with Compliments for information to:-

1) The Chairman, Board of Deans,
2) The Dean, Faculty of Science & Technology,
3) The Chairman, **Ad-hoc Board of Studies in Data Science**,
4) The Director, Board of Examinations and Evaluation,
5) The Director, Board of Students Development,
6) The Director, Department of Information & Communication Technology,
7) The Director, Institute of Distance and Open Learning (IDOL Admin), Vidyanagari.
8) The Deputy Registrar, Admissions, Enrolment, Eligibility & Migration Department (AEM),

Copy forwarded for information and necessary action to :-	
1	The Deputy Registrar, (Admissions, Enrolment, Eligibility and Migration Dept)(AEM), dr@eligi.mu.ac.in
2	The Deputy Registrar, Result unit, Vidyanagari drresults@exam.mu.ac.in
3	The Deputy Registrar, Marks and Certificate Unit,. Vidyanagari dr.verification@mu.ac.in
4	The Deputy Registrar, Appointment Unit, Vidyanagari dr.appointment@exam.mu.ac.in
5	The Deputy Registrar, CAP Unit, Vidyanagari cap.exam@mu.ac.in
6	The Deputy Registrar, College Affiliations & Development Department (CAD), deputyregistrar.uni@gmail.com
7	The Deputy Registrar, PRO, Fort, (Publication Section), Pro@mu.ac.in
8	The Deputy Registrar, Executive Authorities Section (EA) eau120@fort.mu.ac.in He is requested to treat this as action taken report on the concerned resolution adopted by the Academic Council referred to the above circular.
9	The Deputy Registrar, Research Administration & Promotion Cell (RAPC), rapc@mu.ac.in
10	The Deputy Registrar, Academic Appointments & Quality Assurance (AAQA) dy.registrar.tau.fort.mu.ac.in ar.tau@fort.mu.ac.in
11	The Deputy Registrar, College Teachers Approval Unit (CTA), concolsection@gmail.com
12	The Deputy Registrars, Finance & Accounts Section, fort draccounts@fort.mu.ac.in
13	The Deputy Registrar, Election Section, Fort drelection@election.mu.ac.in
14	The Assistant Registrar, Administrative Sub-Campus Thane, thanesubcampus@mu.ac.in
15	The Assistant Registrar, School of Engg. & Applied Sciences, Kalyan, ar.seask@mu.ac.in
16	The Assistant Registrar, Ratnagiri Sub-centre, Ratnagiri, ratnagirisubcentre@gmail.com

Copy for information :-	
1	P.A to Hon'ble Vice-Chancellor, vice-chancellor@mu.ac.in
2	P.A to Pro-Vice-Chancellor pvc@fort.mu.ac.in
3	P.A to Registrar, registrar@fort.mu.ac.in
4	P.A to all Deans of all Faculties
5	P.A to Finance & Account Officers, (F & A.O), camu@accounts.mu.ac.in

1	The Chairman, Board of Deans
2	The Dean, Faculty of Humanities,
3	Chairman, Board of Studies,
4	The Director, Board of Examinations and Evaluation, dboee@exam.mu.ac.in
5	The Director, Board of Students Development, dsd@mu.ac.in@gmail.com DSW direcotr@dsw.mu.ac.in
6	The Director, Department of Information & Communication Technology,
7	The Director, Institute of Distance and Open Learning (IDOL Admin), Vidyanagari, director@idol.mu.ac.in

AC – 12/07/2024

Item No. – 6.1 (N)

As Per NEP 2020

University of Mumbai

Title of the program

M.Sc. (Data Science)

Syllabus for

Semester – Sem.- III & IV
Ref: GR dated
16th May, 2023 for Credit Structure of PG

(With effect from the academic year 2024-25)

University of Mumbai

(As per NEP 2020)

Sr.No.	Heading	Particulars
1	Title of program O:_____ B	M.Sc. (Data Science)
2	Scheme of Examination R: _____	NEP 50% Internal 50% External, Semester End Examination Individual Passing in Internal and External Examination
3	Standards of Passing R: _____	40%
4	Credit Structure R: SP-100A R: SP-100B	Attached herewith
5	Semesters	Sem. III & iV
6	Program Academic Level	6.5
7	Pattern	Semester
8	Status	New
9	To be implemented from Academic Year	2024-25

Sign of the Chairperson
Dr. Mrs. R. Srivaramangai
Ad-hoc BoS (Data Science)

Sign of the Offg. Associate Dean
Dr. Madhav R. Rajwade
Faculty of Science & Technology

Sign of Offg. Dean,
Prof. Shivram S. Garje
Faculty of Science & Technology

Credit Structure of the Program

R: SP-100B

Year	Level	Sem (2yr)	Major				RM	OJT/FP	RP	Cum.Cr.	Degree
			Mandatory			Electives					
2	6.5	Sem III	2*4+2*2+2			4	-	-	RP (607) 4	22	PG Degree after 3-yr UG
			Next Generation Databases (601)	TH	4	Image and Video Analytics (606a) 2TH +2PR **(OR)** Econometrics and Finance (606b) 2TH +2PR **(OR)** Data Science for Agriculture (606c) 2TH+2PR					
			Next Generation Databases Practical (602)	PR	2						
			Text Mining and Natural Language Processing (603)	TH	4						
			Text Mining and Natural Language Processing Practical (604)	PR	2						
			Data Compliance and Auditing(605)	TH	2						
		Sem IV	2*4+2*2			4	-	-	RP (616) 6	22	
			Deep Neural Networks (611)	TH	4	Blockchain Technologies for Data Science (615a) 2TH+2PR **(OR)** Financial Risk Analytics and Management (615b) 2TH+2PR **(OR)** Legal Analytics (615c) 2TH +2PR					
			Deep Neural Networks Practical (612)	PR	2						
			Optimization Methods for Data Science (613)	TH	4						
			Optimization Methods for Data Science Practical (614)	PR	2						
Cum.Cr. For 1 Yr PG Degree			26			8			10	44	
Cum.Cr. For 2 Yr PG Degree			54			16	4	4	10	88	

Semester III

SEMESTER -III

Programme Name: M.Sc. Data Science Semester III	**Course Name:** Next Generation Databases
Total Credits: 04	**Total Marks:** 100
College assessment: 50	**University assessment:** 50

Prerequisites: Basic knowledge about databases and DBMS

Course Outcome:

- Understand the fundamentals of Big Data, its applications in various domains and the technologies in Bigdata ecosystem.
- 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
- Differentiate between ACID and BASE, Understand HDFS File Structure, Map Reduce Framework
- Understand the different types of NoSQL databases like document database, columnar database, and graph database

Course Code	Next Generation Databases	Total Credits
PSDS601	Next Generation Databases	**04**
MODULE I **Unit I** 1. 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. 2. 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. **Unit II** 1. 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) 2. 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,		**02**

MapReduce, aggregate (), MongoDB Document Data Model Approach	
MODULE II **Unit III** 1. 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 2. 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. **Unit IV** 1. 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. **2.** 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	**02**

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

Programme Name: M.Sc. Data Science Semester III **Total Credits:** 02	**Course Name:** Next Generation Databases Practical **Total Marks:** 50 **University assessment:** 50

Prerequisite: Database Fundamentals, Programming Skills

Course Outcome:

- Perform DDL, DML and TCL SQL queries on any RDBMS
- Perform ETL operations on any public cloud platform
- Gain knowledge about bigdata ecosystem and various bigdata technologies
- Gain thorough knowledge about the structural and conceptual difference between different type of NoSQL systems and their implementations.

Course Code	Course Title	Credits
PSDS602	**Next Generation Database Practical**	**02**
Note: - MySQL/ PostgreSQL, Snowflake , Hadoop, MongoDB, Neo4j		
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	

Programme Name: M.Sc. Data Science Semester III **Total Credits:** 04 **College assessment:** 50	**Course Name: Text Mining and Natural Language Processing** **Total Marks:** 100 **University assessment:** 50

Prerequisite:

- Basic understanding of programming (Python preferred)
- Familiarity with machine learning concepts
- Basic statistics and probability

Course Outcome:

- Able to pre-process, clean, and represent text data effectively.
- Implement and evaluate text classification and topic modeling techniques.
- Understand and apply advanced NLP methods

Course Code	Course Title	Total Credits
PSDS603	**Text Mining and Natural Language Processing**	**04**
MODULE I **Unit 1: Introduction to Text Mining and NLP** ● Overview of text mining and NLP ● Applications of NLP in various domains ● Key challenges in text mining (e.g., handling unstructured data, language ambiguity) ● Introduction to Python libraries for NLP (NLTK, spaCy) **Unit 2: Text Pre-processing and Representation** ● Techniques: tokenization, stemming, lemmatization ● Stopwords removal and text normalization ● Text cleaning: handling misspellings, special characters, and case normalization ● Methods: Bag-of-Words (BoW), Term Frequency-Inverse Document Frequency (TF-IDF) ● Word, Document, and sentence embeddings		**02**
MODULE II **Unit 3: Text Classification** ● Overview of text classification tasks ● Naive Bayes, SVM, and Logistic Regression for text classification ● Evaluation metrics for classification ● Introduction to topic modeling, Latent Dirichlet Allocation (LDA) ● Sentiment analysis, Named Entity recognition, Part of Speech tagging **Unit 4: Advanced NLP Techniques and Transformer Models** ● Recurrent Neural Networks (RNNs) ● Long Short-Term Memory (LSTM) ● Implementation using Keras and TensorFlow ● Introduction to transformers (BERT, GPT) ● Pre-trained models and fine-tuning		**02**

Text Books:

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

Reference Books:

1. Neural Network Methods for Natural Language Processing by Yoav Goldberg, 2017, Morgan & Claypool Publishers
2. Foundations of Statistical Natural Language Processing by Christopher D. Manning and Hinrich Schütze, 1999, MIT Press

Programme Name: M.Sc. Data Science Semester III **Total Credits:** 02	**Course Name:** Text mining and Natural Language Processing Practical **Total Marks:** 50 **University assessment:** 50

Course Outcome:

- Understand fundamental concepts of NLP and its processing
- Apply NLP techniques in various application of NLP

Course Code	Course Title	Credits
PSDS604	**Text Mining and Natural Language Processing Practical**	**02**
1	Pre-processiang methods for text data (tokenization, stop-words etc…)	
2	Implement Stemming	
3	Implement Morphological Analysis	
4	Implement N-gram Model	
5	Implement Part-of-Speech tagging	
6	Implement Chucking	
7	Implement Text Summarization	
8	Implement Named Entity Recognition	
9	Implement Sentiment Analysis	
10	Implement One real time NLP Application on a dataset available.	

Programme Name: M.Sc. Data Science Semester III	**Course Name:** Data Compliance and Auditing
Total Credits: 02	**Total Marks:** 50
College assessment: 25	**University assessment:** 25

Prerequisites: Understanding of Relevant Regulations, Basic Understanding of Data Management

Course Outcome:

- Gain a comprehensive understanding of major data protection and privacy regulations such as GDPR, and their implications for businesses.
- Develop the ability to plan, conduct, and document data audits using industry-standard methodologies and tools to ensure data accuracy and compliance.

Course Code	Course Title	Total Credits
PSDS605	**Data Compliance and Auditing**	**02**
MODULE I **Unit 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. **Unit 2: 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. **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. **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		**02**

Security Risk Assessment, Risk Treatment, Residual Risk, Risk Acceptance, Risk Management Feedback Loops etc. **Configuration Reviews:** Introduction to Configuration Management, Configuration Management Requirements-Plan Control, Development of configuration Control Policies, Testing Configuration Management etc.	

Text Books:

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

ELECTIVES

Programme Name: M.Sc. Data Science Semester III	**Course Name:** Image and Video Analytics
Total Credits: 02	**Total Marks:** 50
College assessment: 25	**University assessment:** 25

Prerequisites:

- Basic programming skills in Python.
- Foundational knowledge of data science and machine learning concepts.
- Familiarity with mathematical concepts such as linear algebra, probability, and calculus.

Course Outcome:

By the end of the course, students will be able to:

- Describe the key concepts of image and video analytics.
- Apply image processing techniques to enhance and segment images.
- Extract and describe features from images for various applications.
- Implement motion analysis and object tracking algorithms in videos.
- Utilize video analytics for event recognition and anomaly detection.

Course Code	Course Title	Total Credits
PSDS606a	**Image and Video Analytics**	**02**
MODULE 1: **Unit 1: 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 **Unit 2: 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		**02**

Text Books:

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

Reference Books:

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

Programme Name: M.Sc. Data Science Semester III **Total Credits:** 02	**Course Name:** Image and Video Analytics Practical **Total Marks:** 50 **University assessment:** 50

Prerequisite: Basic understanding of computer science principles, familiarity with programming (Python preferred), and introductory knowledge of signal processing or computer vision.

Course Outcome:

- Capture, manipulate, and format images and video frames using OpenCV.
- Implement and evaluate image enhancement and segmentation methods.
- Utilize algorithms to extract and match features between images.
- Apply transformations and morphological operations for complex image analysis.
- Implement motion detection and object tracking in dynamic video scenes.
- Execute face detection and event recognition for video summarization.

Course Code	**Course Title**	**Credits**
PSDSP606a	**Image and Video Analytics Practical**	**02**
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.
11.	Face Detection Using Haar Cascades: Implement face detection in images and videos using Haar cascade classifiers.
12.	Multi-Object Tracking: Track multiple objects in videos using algorithms like SORT or Deep SORT.
13.	Video Event Detection and Summarization: Detect events and create summarized videos highlighting key activities.

Programme Name: M.Sc. Data Science Semester III **Total Credits:** 02 **College assessment: 25**	**Course Name:** Econometrics and Finance (Elective) **Total Marks: 50** **University assessment: 25**

Prerequisite: Students should be familiar with basic concepts in probability theory and statistical inference

Course Outcome:

- Learners will understand the concepts of Econometrics
- Learners will be able to understand different stages in model development.

Course Code	Course Title	Total Credits
PSDS606b	**Econometrics and Finance**	02
MODULE I **Unit 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. **Unit 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.		**02**

Reference Books:

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

Programme Name: M.Sc. Data Science Semester III **Total Credits:** 02	**Course Name:** Econometrics and Finance Practical **Total Marks:** 50 **University assessment:** 50

Prerequisite: Basic computer skills and Basic understanding of Ms Excel

Course Outcome:

- Learners will learn practical application of Econometrics.
- Learners will be able to use Econometrics in financial modelling

Course Code	Course Title	Credits
PSDSP606b	**Econometrics and Finance Practical**	**02**
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	

Programme Name: M.Sc. Data Science Semester III **Total Credits:** 02 **College assessment:** 25	**Course Name:** Data Science for Agriculture **Total Marks:** 50 **University assessment:** 25

Course Outcome:

- Proficiency in data collection, preprocessing, and analysis for agricultural data.
- Ability to build predictive models and perform EDA specific to agriculture.
- Skills in remote sensing, GIS, and precision agriculture techniques.
- Competence in developing data-driven solutions for agricultural challenges.
- Enhanced understanding of sustainability and environmental impacts in agriculture.

Course Code	Course Title	Total Credits
PSDS606c	**Data Science for Agriculture**	**04**
MODULE I **Unit 1:** 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. **Unit 2:** 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.		**02**

Text Books:

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

Programme Name: M.Sc. Data Science Semester IV **Total Credits:** 02	**Course Name:** Data Science for Agriculture Practical **Total Marks:** 50 **University assessment:** 50

Prerequisites: Foundation in statistics and mathematics, Python or R

Course Outcome:

- Understand and Apply Data Collection Techniques
- Perform Exploratory Data Analysis (EDA)
- Develop Predictive Models
- Implement Precision Agriculture Techniques
- Utilize Remote Sensing and GIS
- Forecast Market Prices and Analyze Supply Chains

Course Code	Course Title	Credits
PSDSP606c	**Data Science for Agriculture Practical**	**02**
Note: - Python, R, data analysis libraries (Pandas, NumPy, Scikit-learn), and data visualization tools (Matplotlib, Seaborn, Tableau).		
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.	

Programme Name: M.Sc. Data Science Semester III	**Course Name: Research Project**
Total Credits: 04	**Total Marks:** 100
College assessment: 50	**University assessment:** 50

Guidelines for Research Project

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 with Certificate is required to submit at the time of Viva Examination

Scheme of Examination for Research Project

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 Viva 2				
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

Semester IV

Programme Name: M.Sc. Data Science Semester IV	**Course Name:** Deep Neural Networks
Total Credits: 04	**Total Marks:** 100
College assessment: 50	**University assessment:** 50

Prerequisite: Linear algebra, foundation of programming and fundamental knowledge of machine learning concepts and basic algorithms

Course Outcome:

- Understand the architectures and workings of various deep learning models
- Implement deep learning solutions for complex problems in computer vision, natural language processing.
- Evaluate and optimize deep learning models using modern techniques and frameworks
- Adapt with cutting-edge research in deep learning to formulate and solve novel research questions.

Course Code	Course Title	Total Credits
PSDS611	**Deep Neural Networks**	**04**
MODULE I Core Deep Learning Architectures **Unit-1: 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 **Unit-2: 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		**02**
MODULE II Modern Deep Learning Techniques **Unit-1: 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 **Unit 2: 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		**02**

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

Programme Name: M.Sc. Data Science Semester IV **Total Credits:** 02	**Course Name:** Deep Neural Networks Practical **Total Marks:** 50 **University assessment:** 50

Prerequisite: Strong skills in Python, including experience with libraries such as NumPy, pandas.

Course Outcome:

- Understand and implement various deep learning architectures such as CNNs, RNNs, LSTMs, Autoencoders
- Apply deep learning techniques to problems in image processing, natural language processing, and sequential data analysis.
- Evaluate and optimize deep learning models using modern techniques such as dropout, batch normalization, and advanced optimizers.
- Apply deep reinforcement learning technique for problem solving.

Course Code	Course Title	Credits
PSDS612	**Deep Neural Networks Practical**	**02**
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	
11	Implement a GAN to generate images.	
12	Implement Deep Q-Network	

Programme Name: M.Sc. Data Science Semester IV **Total Credits:** 04 **College assessment:** 50	**Course Name:** Optimization Methods for Data Science **Total Marks:** 100 **University assessment:** 50

Prerequisite: Basic linear algebra, probability, and knowledge of Python to conduct simulation exercises.

Course Outcome:

- Cast minima/maxima problems into optimization framework.
- Learn efficient computational procedures to solve optimization problems. □

Course Code	Course Title	Total Credits
PSCS613	**Optimization Methods for Data Science**	**04**
MODULE I **Unit 1: Mathematical Foundation** ● Linear algebra and matrices ● Vector space and eigen analysis ● Elementary multivariable calculus ● Optimization and Optimality ● General Formulation of Optimization Problems. **Unit 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		**02**
MODULE II **Unit 3: Unconstrained and Constrained Optimization** ● One-dimensional search methods ● Gradient-based methods ● Conjugate direction and quasi-Newton methods ● Lagrange theorem ● FONC, SONC, and SOSC conditions **Unit 4: Non-Linear Problem and Nature-Inspired Algorithms** ● Non-linear constrained optimization models ● 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.		**02**

Text Books:

1. Optimization Techniques and Applications with Examples by Xin-She Yang Wiley 3rd 2018
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. Nonlinear Programming by Dimitri Bertsekas, Athena Scientific, 1999

Programme Name: M.Sc. Data Science Semester IV **Total Credits:** 02	**Course Name:** Optimization Methods for Data Science Practical **Total Marks:** 50 **University assessment:** 50

Prerequisite: Basic computer skills and Basic understanding of programming language (python)

Course Outcome:

- Be able to model minima/maxima problems as optimization problems.
- Use Python to implement important optimization methods.

Course Code	Course Title	Credits
PSDSP614	Practical on Optimization Methods for Data Science	**02**
Note: Working on **- https://jupyter.org/try-jupyter/**		
1.	Matrix operation and Differentiation of vector and matrix.	
2.	Integration of a vector and matrix	
3.	Simplex algorithm and Duality	
4.	Implementation of Newton's method	
5.	Implementation of Secant method	
6.	Implementation of Lagrange multiplier method	
7.	Implementation of KKT theorem	
8.	Implementation of BFGS method	
9.	Particle Swarm Optimization Algorithm	
10.	Flower Pollination Algorithm	

ELECTIVES

Programme Name: M.Sc. Data Science Semester IV **Total Credits:** 02 **College assessment:** 25	**Course Name:** Blockchain Technologies for Data Science **Total Marks:** 50 **University assessment:** 25

Prerequisite:
Basic understanding of data science concepts and proficiency in programming languages such as Python or JavaScript.

Course Outcome:

- Students will gain a thorough understanding of blockchain technology, including its architecture, key concepts, and different platforms.
- Students will acquire hands-on experience in developing and deploying smart contracts, setting up blockchain networks, and integrating blockchain with data science applications.
- Students will be able to apply blockchain technology to ensure data integrity, security, and provenance in data science projects.
- Students will develop the ability to create innovative blockchain-based solutions for real-world data science problems, including decentralized storage, data marketplaces, and secure data sharing.

Course Code	Course Title	Total Credits
PSDS615a	**Blockchain Technologies for Data Science**	**02**
MODULE I **Unit 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 **Unit 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, Building a Data Marketplace, Hyperledger Fabric Chaincode, Blockchain-based Voting System, Blockchain and IoT Integration	**02**

Text Books:

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.

Reference Books:

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.

Programme Name: M.Sc. Data Science Semester IV **Total Credits:** 02	**Course Name:** Blockchain Technologies for Data Science Practical **Total Marks:** 50 **University assessment:** 50

Prerequisite:
Basic programming knowledge and command-line skills, Understanding of blockchain concepts and decentralized systems, Familiarity with data management and machine learning basics, Knowledge of security concepts and IoT principles.

Course Outcome:

- Ability to set up blockchain development environments, write and deploy smart contracts, and manage Hyperledger Fabric networks.
- Proficiency in using IPFS for decentralized storage and integrating blockchain to ensure data integrity in machine learning models.
- Capability to implement blockchain-based data provenance systems and develop decentralized data marketplaces.
- Skills to deploy chaincode on Hyperledger Fabric, create secure blockchain-based voting systems, and integrate blockchain with IoT for secure data management.

Course Code	Course Title	Credits
PSDSP615a	**Blockchain Technologies for Data Science Practical**	**02**
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	

Programme Name: M.Sc. Data Science Semester IV **Total Credits:** 02 **College assessment:** 25	**Course Name:** Financial Risk Analytics **Total Marks:** 50 **University assessment:** 25

Prerequisite: Understanding concepts of Mathematics and Statistics

Course Outcome:

- To address the challenges associated with financial risk (market risk, credit risk and operation risk) through quantitative models and statistical methods.
- To provide an overview of data analytics application under financial risk analytics context.
- To gain the knowledge of handling financial risks using data analytics framework.
- To use appropriate risk measures in financial domain.

Course Code	Course Title	Total Credits
PSDS615b	**Financial Risk Analytics**	**04**
MODULE I **Unit-I** **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 **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 **Introduction to Financial Markets.** Data from FRED, Yahoo, and other sources. Empirical characteristics of economic and financial time series. Boostrapping confidence intervals. **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. **Unit-II** **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. **Measuring Volatility.** Measuring volatility. GARCH estimation. GARCH simulation. Measuring Value at Risk (VaR) and Expected Shortfall (ES). **Portfolio Optimization.** Combining risk management with portfolio allocations. Optimizing allocations. Simulating the efficient frontier. **Aggregating Enterprise Risks.** Enterprise risk management analytics and application.		**02**

Text Books:

1. Financial Risk Analytics: Measurement, Management and Examples in R Kindle Edition by R. K. Arora (Author), Prerna Lal, Willey,20th April 2022
2. Analytics in Finance and Risk Management by Nga Thi Hong Nguyen, Shivani Agarwal, Ewa Ziemba Released December 2023, Publisher(s): CRC Press, ISBN: 9781003808664

Reference Books:

1. Financial Engineering Analytics: A Practice Manual Using R, *William G. Foote,2018-01-09*
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

Programme Name: M.Sc. Data Science Semester III **Total Credits:** 02	**Course Name:** Financial Risk Analytics Practical **Total Marks:** 50 **University assessment:** 50

Prerequisite: Familiarity with libraries such as NumPy, pandas, scikit-learn, and statsmodels (for Python) or dplyr and ggplot2 (for R)

Course Outcome:

- Apply the knowledge of data analytics techniques to handle the risk in financial sector.
- Analyze the existing and potential data analytic solutions in financial risk management.

Course Code	Course Title	Credits
PSDSP615b	**Financial Risk Analytics Practical**	**02**
Note: - R-4.4.0 for Windows		
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: ● 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: 1. Measure risks using historical and parametric approaches 2. Interpret results relative to business decisions 3. Visualize market risk	

5	Credit Risk: ● 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	Operational Risk: ● Define frequency and Severity ● Calculate risk of loss ● Calculate potential loss ● Fire losses ● Estimating the extremes
7	Measuring Volatility: ● Use a fix for volatility clustering ● Fit AR-GARCH models ● Simulate volatility from the AR-GARCH model ● Measure the risks of various exposures
8	Portfolio Analytics: ● Portfolio Optimization. ● Combine risk management with portfolio allocations. ● Optimizing allocations.
9	Perform Risk analysis using Monte Carlo Simulations.
10	Build an App The application will contain four architectural layers. - 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 - User Interface (UI) 1. Slide bars for user to input range of parameters 2. Plots to display results 3. Text to report results - Server 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

Programme Name: M.Sc. Data Science Semester IV	**Course Name:** Legal Analytics
Total Credits: 02	**Total Marks:** 50
College assessment: 25	**University assessment:** 25

Prerequisite: Basic statistical concepts including probability, distributions, hypothesis testing, and regression analysis, Basics of programming (Python/R), data structures, and algorithms.

Course Outcome:

- Gain a comprehensive understanding of various legal data sources, including case law, statutes, regulations, and legal documents.
- Understand the ethical considerations and legal regulations surrounding data privacy in the context of legal analytics.
- Develop skills in cleaning, preprocessing, and exploring legal datasets to identify patterns and trends.
- Machine Learning for Legal Analytics
- Apply text mining and NLP techniques to legal texts to extract meaningful insights.
- Use advanced NLP techniques such as topic modeling and deep learning for complex legal text analysis.
- Gain hands-on experience with legal analytics tools and platforms, integrating analytics into law practice management.

Course Code	Course Title	Total Credits
PSDS615c	**Legal Analytics**	**02**
MODULE I **Unit 1** 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) Data Privacy and Ethics: Ethical considerations in legal analytics, Privacy laws and regulations (e.g., GDPR, CCPA), Data anonymization and security Data Cleaning and Preparation: Handling missing data, Data normalization and transformation, Feature engineering for legal datasets. 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 Text Mining and Natural Language Processing (NLP): NLP techniques for legal text: tokenization, stemming, lemmatization, Sentiment analysis and named entity recognition in legal documents **Unit 2** 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		**02**

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	

Text Books:

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

Programme Name: M.Sc. Data Science Semester IV **Total Credits:** 02	**Course Name:** Legal Analytics Practical **Total Marks:** 50 **University assessment:** 50

Prerequisite: Python and relevant libraries (NumPy, pandas, scikit-learn, NLTK, spaCy), Machine learning techniques applicable to legal data

Course Outcome:

1. Proficiency in using databases and SQL for managing large datasets.
2. Use data visualization tools to represent legal data and insights effectively.
3. Evaluate and interpret the performance of machine learning models in the context of legal data.
4. Utilize natural language processing (NLP) techniques to analyze legal texts, such as court opinions, contracts, and statutes.
5. Develop predictive models to forecast legal trends and outcomes (e.g., litigation risk, case duration).
6. Apply data analytics to support legal decision-making processes.

Course Code	Course Title	Credits
PSDSP615c	Legal Analytics Practical	**02**
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	Develop a machine learning model to predict the outcome of court cases. • 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	Apply clustering algorithms to group similar legal documents. • Preprocess the text data. • Use TF-IDF to convert text to numerical features. • Apply k-means clustering and visualize the clusters
7	Use topic modeling to identify themes in legal documents. • 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 legal documents. • Load and preprocess the legal text data. • Apply spaCy's NER model to extract entities. • Visualize the extracted entities and analyze their occurrences.
9	Extract and analyze key clauses from a set of contracts. • 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	Create a QA system that can answer questions using a dataset of legal documents. • 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.

Programme Name: M.Sc. Data Science Semester IV	**Course Name:** Research Project
Total Credits: 06	**Total Marks:** 150
College assessment: 75	**University assessment:** 75

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. The project documentation needs to be signed by the teacher in charge and head of the Department.

Student should also attach the certified copy of the internal evaluation report (Appendix III) at the time of Project evaluation and viva as part of the University examination.

Scheme of Examination for Research Project

Internal Examination

B) Continuous Internal Evaluation:

Method		**Marks**		
Internal Viva 1		**40**		
Methodology	**Implementation**	**Total**		
20	20	40		
Internal Viva 2		**35**		
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/Model**	**Results**	**Documents**	**Viva**	**Total**
05	05	10	20	20	05	10	75

EVALUATION SCHEME

A. Evaluation for Mandatory Theory Courses (4 Credit Courses)

I. Internal Evaluation for Mandatory Theory Courses – 50 Marks

A. 40 marks (Any one of the following):

a. Written Test or
b. SWAYAM (Advanced Course) of minimum 20 hours and certification exam completed or
c. NPTEL (Advanced Course) of minimum 20 hours and certification exam completed or
d. Valid International Certifications (Prometric, Pearson, Certiport, Coursera, Udemy and the like)
e. One certification marks shall be awarded one course only. For four courses, the students will have to complete four certifications.

B. 10 marks: Class participation, Question answer sessions during lectures, Discussions

II. External Examination for Mandatory Theory Courses – 50 Marks

- Duration: **2 Hours**
- Theory question paper pattern:

	All questions are compulsory.		
Question	**Based on**	**Options**	**Marks**
Q.1	Unit I	*Any 2 out of 4*	10
Q.2	Unit II	*Any 2 out of 4*	10
Q.3	Unit III	*Any 2 out of 4*	10
Q.4	Unit IV	*Any 2 out of 4*	10
Q.5	Unit I, II, III & IV	*Any 2 out of 4*	10

B. Evaluation for Elective Theory Courses (2 Credit Courses)

I. Internal Evaluation for Elective Theory Courses – 25 Marks

A. 15 marks (Any one of the following):

a. Written Test or
b. SWAYAM (Advanced Course) of minimum 20 hours and certification exam completed or
c. NPTEL (Advanced Course) of minimum 20 hours and certification exam completed or
d. Valid International Certifications (Prometric, Pearson, Certiport, Coursera, Udemy and the like)
e. One certification mark shall be awarded one course only. For four courses, the students will have to complete four certifications.

B. 10 marks: Class participation, Question answer sessions during lectures, Discussions

II. External Examination for Elective Theory Courses – 25 Marks

- Duration: **1 Hour**
- Theory question paper pattern:

	All questions are compulsory.		
Question	**Based on**	**Options**	**Marks**
Q.1	Unit I	*Any 2 out of 4*	10
Q.2	Unit II	*Any 2 out of 4*	10
Q.3	Unit I & II	*Any 1 out of 2*	5

C. Evaluation for Mandatory & Elective Practical Courses (2 Credit Courses)

- Each Practical Course carries 50 Marks
  - **40 marks + 05 marks (journal) + 05 marks (viva)**
- 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**

Sign of the Chairperson
Dr. Mrs. R. Srivaramangai
Ad-hoc BoS (Data Science)

Sign of the Offg. Associate Dean
Dr. Madhav R. Rajwade
Faculty of Science & Technology

Sign of Offg. Dean,
Prof. Shivram S. Garje
Faculty of Science & Technology