SCHEME OF INSTRUCTION FOR M.Sc. DATA SCIENCE (2022-24)

II YEAR (2023-24) OF 2022-24 BATCH (CBCS)					
SEMESTER-I			**SEMESTER-II**		
THEORY			**THEORY**		
1	MDS20101	Communicative Competence **(AECC-1)**	1	MDS20201	Human Values & Professional Ethics **(AECC-2)**
2	MDS20102	Technical Seminar and Report Writing **(SEC-1)**	2	MDS20202	Industrial Practices **(SEC-2)**
			3	MDS22203	Multivariate Analysis & Stochastic Process **(GE- InterDisciplinary-1)**
CORE			**CORE**		
3	MDS20103	Mathematics for Data Science **(core 1)**	4	MDS22204	Elective-I **(DSE-1)**
4	MDS22104	Artificial Intelligence **(core 2)**		MDS22204A	Cloud Computing
5	MDS20105	Statistics & Probability **(core 3)**		MDS22204B	Social Media Analytics
6	MDS22106	Python for Data Science **(core 4)**		MDS22204C	Data security and Privacy
7	MDS20107	Advanced Databases **(core 5)**	5	MDS20205	Regression Analysis & Statistical Inference **(core 6)**
			6	MDS22206	Big Data Analytics through Hadoop **(core 7)**
			7	MDS22207	Machine Learning **(core 8)**
PRACTICALS			**PRACTICALS**		
8	MDS22108	Data Visualization Lab **(Lab-I)**	8	MDS20208	Data Analytics Lab **(Lab-IV)**
9	MDS22109	Python for Data Science Lab **(Lab-II)**	9	MDS22209	Big Data Analytics through Hadoop **(Lab-V)**
10	MDS20110	Advanced Databases Lab **(Lab-III)**	10	MDS22210	Machine Learning Lab **(Lab-VI)**

II YEAR (2023-24) OF 2022-24 BATCH (CBCS)					
SEMESTER-III				**SEMESTER-IV**	
THEORY				**THEORY**	
1	MDS21301	Soft Skills (Interview Skills, Group discussion) **(AECC-3)**	1	MDS21401	Major Project
2	MDS21302	Mini Project **(SEC-3)**			
CORE					
3	MDS23303	Time Series and Forecasting Techniques **(core 9)**			
4	MDS23304	Elective-II **(DSE-II)**			
	MDS23304 A	Big Data Security			
	MDS23304 B	Reinforcement Learning			
	MDS23304 C	Scalable Architectures For ML Applications			
5	MDS21305	Internet of Things **(core 10)**			
6	MDS23306	Deep Learning**(core 11)**			
7	MDS21307	Natural Language Processing **(core 12)**			
PRACTICALS					
8	MDS21308	Internet of Things**(Lab-VII)**			
9	MDS23309	Deep Learning Lab**(Lab-VIII)**			
10	MDS21310	Natural Language Processing Lab **(Lab-IX)**			

YEAR-WISE AND SEMESTER-WISE DISTRIBUTION OF SUBJECTS M.Sc. DATA SCIENCE
THIRD SEMESTER
ACADEMIC YEAR 2023-24 OF 2022-24 BATCH (CBCS)

Sl. No.	Subject Code	Title of the Subject	Hours /Week	Duration of Exam (hrs.)	Marks			Credits
					Internal	External	Total	
1	*MDS21301	*Soft Skills**(AECC-3)**	2	-	50	-	50	2*
2	#MDS21302	Mini Project **(SEC-3)**	2	-	25	50	75	2
CORE								
3	MDS23303	Time Series & Forecasting Techniques **(Core-9)**	4	3	40	60	100	4
4	MDS23304 MDS23304A MDS23304B MDS23304C	Elective-II **(DSE-II)** Big Data Security Reinforcement Learning Scalable Architectures For ML Applications	4	3	40	60	100	4
5	MDS21305	Internet of Things **(Core-10)**	4	3	40	60	100	4
6	MDS23306	Deep Learning **(Core-11)**	4	3	40	60	100	4
7	MDS21307	Natural Language Processing **(Core-12)**	4	3	40	60	100	4
PRACTICALS								
7	MDS21308	Internet of Things Lab **(Lab-VII)**	3	3	40	60	100	2
8	MDS23309	Deep Learning Lab (**Lab-VIII)**	3	3	40	60	100	2
9	MDS21310	Natural Language Processing Lab **(Lab-IX)**	3	3	40	60	100	2
Total			**33**		**395**	**530**	**925**	**28**

*** Ability Enhancement Compulsory Course (AECC) is excluded for SGPA and CGPA calculation, but it is compulsory to pass in the examination.**
#In the third Semester students will take one month summer project of 2 credits

CIA components for Internal marks of theory paper: i) Assignment: 3M ii) Attendance: 4M iii)Group Discussion OR Presentation OR Case Study OR Computer Based Test : 5M iv) Mid Exam 1 : 12.5M v) Mid Exam 2: 12.5M vi) Seminar or Viva Voce : 3M
CIA components for Practical Internal marks: i) Attendance: 5M ii) Observation/record book: 5M iii) Practical skill acquired: 10M iv) Pre final: 20M.
Remedial class/Communication skills/ Mentoring will be given 3 extra classes in a week.

SOFT SKILLS

Credits:2 **Semester:III**
Subject Code: MDS21301 **No. of lecture hours : 30**

Objective:

- To equip the student with the adequate soft skills required for any organization. And, skills required enhancing his/her career.

Course Outcome:

- The student will be able to develop effective communication skills, presentation skills, inter- personal skills, team management skills, and leadership skills.

UNIT – I **6Hrs**

Goal Setting

- Meaning of goal and goal setting short. 2
- Medium and long term goal setting 2
- Importance of goal setting-Choices/Selection of setting goals-Steps for goal
- setting-SMARTgoals 2

UNIT-II **6Hrs**

Time Management

- What and why of time management 2
- Necessity and benefits of time management-tools of time management 2
- How to manage time wisely 2

UNIT-III **6Hrs**

Etiquettes

- Get the first impression well. 1
- Greet others & introduce yourself. 1
- Body language- speakwell 2
- Dressing sense- appeals to others 2

UNIT-IV **6Hrs**

Group Discussion Skills

- Leadership Skills, InterpersonalSkills. 1
- Persuasive Skills, Problem SolvingSkills. 2
- ConceptualizationSkills. 1
- Initiating the discussion, listening to others point of view. 2

UNIT-V **6Hrs**

Interview Skills

- Creating first impression in an interview, walk up to interview room, 2
- How to approach the interview members, sitting posture in the interview room,
- body language 4

ESSENTIAL READING

1. Atkinson Frank.2011. Get More Out of Your Day.India:Vikas Books Private Limited. (Unit I in Pg 17-29, Unit II in Pg 5-16 and Pg 73 –118)
2. Butterfield Jeff. Soft Skills for Everyone. India:Cengage Learning. (Unit III in Pg 177-204 and Pg 336 -339, Unit IV in Pg 357-373)
3. Philips R. Hunsaker. Training in Interpersonal Skills. New Delhi:McGraw Hill. (Unit V in Pg 176-214)

SUGGESTED READING

1. Time Management. USA:Harvard Business School Press.Boston.
2. Padhy Kishore C and madhuchhanda. 2008.A to Z of Interview. India:Himalaya PublishingHouse.

MINI PROJECT

Credits: 2 **Semester:III**
Subjectcode: MDS21302 **No. of lecture hours:30**

Objective:

- The objective of mini project work is to develop quality analytical solutions. The primary emphasis of the project work is to understand and gain the knowledge of the principles of data science and analytics practices.

Course Outcome:

- The program prepares the students to take up positions as Systems Analysts, Systems Designers, Data scientist, Programmers and Project Managers in any field related to data science and analytics.

The main aim of this mini project is to check the learning skills of students in Python, MachineLearningandDataMiningetc.,whichiscoveredinIandIIsemesters.Thestudents areexpectedtodoaMiniprojectinthesummervacationfollowingthe2ndsemester.Project must be carried out by each student individually in a period of 30 days of duration. Students should submit a synopsis at the end of 2nd semester in consultation with the Project Guide. The synopsis should consist of definition of the problem, scope of the problem and plan of action.

Students are supposed to take this project work very seriously, as these efforts will be the showcase of their skill and knowledge in Python, Machine Learning and Data Mining etc. The mini project should be genuine and original in nature and should not be copied from anywhere else. If found copied, then it will be canceled automatically.

After completion of project, in the 3rd semester students are required to present a Project Seminar covering the aspects of Data Collection, Data Visualization, Labeling, Data Selection, Data Pre-processing, Data transformation, Model Training, Model evaluation, Model testing and accuracy are required to show in the project work.

A committee consisting of two or three faculty members of the department along with a guide will evaluate the project and award internal marks.

TIME SERIES ANALYSIS AND FORECASTING TECHNIQUES

Credits: 4 **Semester: III**
Subject code: MDS23303 **No. of lecture hours: 60**

Objectives:

- To understand the differences between cross-sections and time series and specific economic problems which occur while working with data of these types.

Course Outcomes:

CO1: Identifying linear, quadratic, Gompertz and Logistic models where appropriate and describe models for seasonal variation. Also explains the methods used to study cyclic components.

CO2: Estimate seasonal effects of time-series data by using Winten's method, Brown's, Box – Jenkin's three-parameter exponential smoothing method.

CO3: To utilize AR, ARIMA models for time series data and to forecast the data using these models.

CO4: To interpret SARIMA model and criterion used to study them.

CO5: To explain and verify mathematical considerations for analysing time series, including concepts of stationarity, autocovariance, autocorrelation.

UNIT-I **12Hrs**

Growth Models:

- Modified Exponential curve, Gompertz curve, Logistic curve, and their fitting. 5

Measurement of cyclic Variations:

- Harmonic analysis, Auto – regression series – Markoff and Yule's series. 5
- Periodogram and correlogram analysis. 2

UNIT-II **12Hrs**

Measurement of irregular Component:

- Variant difference method. 2

Exponential Smoothing Methods:

- Trend adjusted exponential smoothing. Double and triple exponential smoothing. 5
- Win ten's method. Brown's one parameter adaptive method. 1
- Box – Jenkin's three parameter smoothing, 2
- Harrison's Harmonic smoothing methods and Tracking signal. 2

UNIT-III **12Hrs**

- Box – Jenkin's Time Series Methods: Moving average, Autoregressive (AR). 4
- ARMA and AR integrated MA (ARIMA) models. 4
- Estimation of ARIMA model parameters forecasting with ARIMA models. 4

UNIT-IV **12Hrs**

- SARIMA (Seasonal Autoregressive Integrated Moving Average method). 2
- Akaike's Information Criterion (AIC) 2

- Bayesian Information Criterion (BIC) 2
- Root Mean Square Error (RMSE) 2
- Diagnostic checking of models. 2
- Analysis of residuals. 2

UNIT-V **12Hrs**

- Stationary time series. Autocovariance and autocorrelation functions. 2
- Regression models for general time series data. 2
- Detecting autocorrelation – Durbin – Watson test. 2
- Estimating the parameters in time series regression models. 2
- ARCH and GARCH modeling of non – linear time series data. 4

ESSENTIAL READING

1. Box, G.E.P and Jenkins, GM: Time Series Analysis Forecasting and Control, Holden Day, San Francisco.
2. Anderson, T.W.: The Statistical Analysis of Time series, John wiley, New York.
3. Thomopouls, N.T.: Applied Forecasting methods. EngleWood cliffs, N.J, Prentice Hall.

SUGGESTED READING

1. Wheel Wishart S.C. and S. Makridaks: Forecasting Methods for Management, 3rd Edition, John Wiley, New York.
2. Sullivan, William G. and Wayne Claycambe. W: Fundamentals of Forecasting, Prentice Hall.

BIG DATA SECURITY

Credits : 4 **Semester:III**
Subjectcode : MDS23304A **No. of lecture hours : 60**
Prerequisites : Big Data Analytics

Objectives:

- Introduction to big privacy, ethics and security
- Learn about steps to secure big data and challenges
- Understand hadoop kerberos security implementation & configuration.
- Learn how to secure different Hadoop ecosystem components

Course Outcomes:

CO1: Learn the fundamentals of big data security.
CO2: To explore the steps to secure big data.
CO3: Understand the different hadoop security design.
CO4: Implementing security for Hadoop ecosystem components.
CO5: Understand integrating Hadoop with Enterprise Security Systems and securing sensitive data in hadoop

UNIT – I **12Hrs**

Big Data Privacy, Ethics And Security

- Privacy, Re Identification of Anonymous People, 3
- Why is Big Data Privacy self regulating? 3
- Ethics, Ownership, Ethical Guidelines 3
- Big Data Security, Organizational Security. 3

UNIT – II **12Hrs**

Security, Compliance, Auditing, And Protection

- Steps to secure big data, Classifying Data 4
- Protecting, Big Data Compliance Intellectual Property Challenge, 4
- Research Questions in Cloud Security, Open Problems. 4

UNIT – III **12Hrs**

Hadoop Security Design

- Kerberos, Default Hadoop Model without security, 6
- Hadoop Kerberos Security Implementation & Configuration. 6

UNIT – IV **12Hrs**

Hadoop Ecosystem Security

- Configuring Kerberos for Hadoop ecosystem components – Pig, Hive, Oozie, Flume, HBase, Sqoop. 12

UNIT – V **12Hrs**

Data Security & Event Logging

- Integrating Hadoop with Enterprise Security Systems, 4
- Securing Sensitive Data in Hadoop, 4
- SIEM system, Setting up audit logging in hadoop cluster. 4

ESSENTIAL READING:

1. Mark Van Rijmenam, “Think Bigger: Developing a Successful Big Data Strategy for Your Business”, Amazon, 1 edition, 2014
2. Frank Ohlhorst John Wiley & Sons, “Big Data Analytics: Turning Big Data into Big Money”, John Wiley & Sons, 2013.
3. Sherif Sakr, “Large Scale and Big Data: Processing and Management”, CRC Press, 2014
4. Ben Spivey, Joey Echeverria, “Hadoop Security Protecting Your Big Data Problem”, O’Reilly Media, 2015

REINFORCEMENT LEARNING

Credits : 4 **Semester:III**
Subject code : MDS23304B **No. of lecture hours : 60**
Prerequisites : Machine Learning

Objectives:
Knowledge on fundamentals of reinforcement learning and the methods used to create agents that can solve a variety of complex tasks.

Course Outcomes:

CO1: Understand basics of Monte-Carlo Method.
CO2: Understand RL model free methods.
CO3: Understand mathematical trick that improves the performance of Temporal Difference
CO4: Acquire knowledge in finding the value of a state or an action when similar circumstances occur.
CO5: Analyze meaningful information from digital images, videos and other visual inputs.

UNIT – I **12Hrs**

Monte-Carlo Methods

- Monte-Carlo methods: policy evaluation, 4
- Rollouts, on policy and off-policy learning, 4
- Importance sampling 4

UNIT – II **12Hrs**

Temporal Difference Learning

- Temporal Difference learning: TD prediction, 3
- Optimality of TD (0), SARSA, 3
- Q-learning, Games and after states, 3
- Maximization Bias and Double Learning. 3

UNIT – III **12Hrs**

EligibilityTraces

- Eligibility traces: n-step TD prediction, 4
- TD (lambda), forward and backward views, 4
- Q(lambda), SARSA(lambda), Replacing traces and accumulating traces. 4

UNIT – IV **12Hrs**

Function Approximation

- Function Approximation: Value prediction, 4
- gradient descent methods, linear function approximation, Control algorithms, 4
  Fitted Iterative Methods, Deep Q-learning. 4

UNIT – V **12Hrs**

Computer Vision

- Policy Gradient methods: non-associative learning - REINFORCE algorithm, 6
- Exact gradient methods, estimating gradients, approximate policy gradient algorithms, actor-critic methods, Asynchronous Advantage Actor-Critic.
- Hierarchical RL: MAXQ framework, Options framework, HAM framework, Option discovery algorithms. 4
- Case studies: Elevator dispatching, Samuel's checker player, TD- gammon, Acrobot, Helicopter piloting, Alpha Go 2

ESSENTIAL READING:

1. R. S. Sutton and A. G. Barto. Reinforcement Learning - An Introduction. MIT Press. 2nd Edition. 2018
2. Video Lectures by Prof. David Silver
3. Video Lectures by Prof. B.Ravindran

SCALABLE ARCHITECTURES FOR ML APPLICATIONS

Credits : 4 **Semester:III**
Subjectcode : MDS23304C **No. of lecture hours : 60**
Prerequisites : Machine Learning

Objectives:

- Students will be able to learn application and building Scalable Machine Learning
- Understand the students with Hadoop, SMACK Stack and also Message Services.
- Be able to select the appropriate architecture for enterprise architectures based on the size, scale and applications used in the enterprise.

Course Outcomes:

CO1: Understand the basic concepts of Scalable Machine Learning
CO2: To become a data scientist work in some development environment tailored for statistics and Machine Learning
CO3: Obtain expertise to turn actionable insights and Fast Data Applications into innovative methods to solve real-world problems
CO4: Demonstrate the graph algorithms and live streaming data in Spark
CO5: To impart knowledge on Kubernetes and batch processing

UNIT – I **12Hrs**

- Introduction to Scalable Machine Learning, 4
- Some Machine Learning Background Algorithms for Large scale Learning, 4
  Overview of Hadoop and Current Big Data Systems 4

UNIT – II **12Hrs**

- How Programming for Data Flow Differs, 3
- Basic Spark, Working with Vectors and Matrices in Spark, 3
- Brief tour of Spark ML, 3
- beyond parallelization, Practical Big Data 3

UNIT – III **12Hrs**

- Anatomy of Fast Data Applications, 3
- SMACK Stack – Functional Decomposition, 3
- Message Backbone- Understanding messaging requirements, Data 3
  ingestion, Fast data& low latency, Message Delivery Semantics, 3
  Distributing Messages

UNIT – IV **12Hrs**

- Compute Engines- Micro Batch Processing, 3
- One-at-a time Processing, Choice of processing engine, 3
- Storage as the Fast Data Borders, 3
- The message backbone as Transition Point 3

UNIT – V **12Hrs**

- Sharing stateful streaming state, Data Driven Micro-services, 3
- State and Micro-services. Deployment environments for Fast Data Applications, Application containerization, 3
- resource scheduling, Apache Mesos, Kubernetes, Cloud Deployments. 3

ESSENTIAL READING:

1. Designing Fast Data Application Architectures by Gerard Maas, Stavros Kontopoulos, Sean Glover , Publisher: O'Reilly Media, Inc., June 2018
2. Spark- The definitive Guide by Bill Chambers & Matei Zaharia, O'Reilly Media, Inc., June 2019

INTERNET OF THINGS

Credits:4 **Semester:III**
Subjectcode : MDS21305 **No. of lecture hours:60**

Objectives:

- To introduce the terminology, technology, and its applications
- To introduce the concept of M2M (machine to machine) with necessary protocols
- To introduce the deployment of IOT design methodology
- To introduce the Raspberry PI platform, that is widely used in IoT applications.
- To introduce the implementation of web-based services on IoT devices

Course Outcomes:

CO1: **Identify** the importance of IOT and its applications.
CO2: **Differentiate** between IOT and M2M, SDN and NFV
CO3: **Apply** IOT design methodology.
CO4: **Understand** building of IOT devices and Raspberry PI.
CO5: **Explain** working of application of IOT.

UNIT -I **12 Hrs**

Introduction and Concepts

- Introduction to Internet of Things –Definition and Characteristics of IoT 3
- Physical Design of IoT, Logical Design of IoT 3
- IoT Enabling Technologies 3
- IoT Levels and Deployment Templates, Domain Specific IoTs – Home Automation Cities, Environment, Agriculture, Industry, health, and Lifestyle 3

UNIT – II **12 Hrs**

IoT and M2M

- IoT and M2M- Introduction to M2M, Difference between IoTandM2M 3
- SDN and NFV for IoT 3

IoT System Management with Netconf-Yang

- Need for IoT Systems Management, SNMP 2
- Network Operator requirements, NETCONF,YANG 2
- IoT Systems Management with NETCONF-YANG 2

UNIT -III **12 Hrs**

IoT Design Methodology

- Design methodology 2
- IOTlevel1,2,3 2

- IOTlevel4,5,6 2
- IOTlevel7,8,9,10 3
- Weather monitoring system case study 3

UNIT -IV **12Hrs**

IoT Physical Devices and Endpoints

- Building blocks of IoT device 2
- Raspberry Pi, About the Board, Linux on Raspberry Pi, RaspberryPi Interfaces 3
- Programming RaspberryPi with Python 5
- Other IoT devices 2

UNIT -V **12Hrs**

IoT Physical Servers and Cloud Offerings

- Introduction to Cloud Storage models andCommunication API 3
- WAMP-AutoBahn for IoT, Xively CloudforIoT 3
- Python web application framework-Django 3
- Case study on IOT applications 3

ESSENTIAL READING

1. Bahga, Arshdeep and Madisetti, Vijay. 2015. **Internet of Things - A Hands-on Approach.** Universities Press ISBN:9788173719547
2. Richardson, Matt and Wallace, Shawn. 2014. **Getting Started with RaspberryPi.** O'Reilly (SPD). ISBN: 9789350239759
3. Internet of Things Projects with Esp32: Build exciting and powerful IoT projects, 1st Edition, by AgusKurniawan

DEEP LEARNING

Credits: 4 **Semester:III**
Subjectcode: MDS23306 **No. of lecture hours:60**
Prerequisites: Machine Learning

Objectives:

- To build the foundation of neural networks.
- Explore advanced topics of neural networks.
- To design and analyze various algorithms and techniques with a modern outlook.
- To enable the students, develop successful deep learning projects.

Course Outcomes:

CO1: Learn the fundamental principles of deep learning.
CO2: Identify the deep learning algorithms for various types of learning tasks in various domains.
CO3: To explore Deep learning techniques and various feature extraction strategies.
CO4: To mathematically understand the deep learning approaches and paradigms.
CO5: Implement deep learning algorithms and solve real-world problems.

UNIT – I	**12Hrs**
Introduction:	
● Historical context and motivation for deep learning	2
● Supervised Learning Algorithms	2
● Unsupervised Learning Algorithms	2
● Optimizing logistic classifier using gradientdescent	2
● Stochastic gradient descent, momentum, and adaptivesub-gradient method.	4

UNIT – II	**12Hrs**
Neural Networks:	
● Feedforward neural networks	2
● Deep Networks	2
● ArchitectureDesign	2
● Regularizing a Deep network	2
● Model exploration and hyperparameter tuning.	4

UNIT – III	**12Hrs**
Convolution Neural Networks:	
● Introduction to convolution neural networks	2
● Motivation,Pooling	2
● Convolution and Pooling as an Infinitely strong prior	2

- Variants of the basic Convolution function 2
- Structured outputs, Efficient ConvolutionAlgorithms 4

UNIT – IV **12Hrs**

Sequence Modeling : Recurrent Nets and Recursive Nets:

- Unfolding computational graphs, recurrent neural networks(RNNs) 2
- Bidirectional RNNs 2
- Encoder-Decoder sequence to sequence architectures 2
- Deep Recurrent networks 2
- Recursive Networks 4

UNIT – V **12Hrs**

Auto encoders:

- Under complete auto encoders 2
- Regularized auto encoders, sparse autoencoders, denoising auto encoders 2
- Representational power, layer, size, and depth of auto encoders 2
- Stochastic encoders and decoders 2
- Application of auto encoders 4

ESSENTIAL READING:

1. Deep Learning by Ian Goodfellow, YoshuaBengio. Aaron Courville. The MIT Press, 2016

SUGGESTED READING:

1. Deep Learning, A Practical Approach by Rajiv Chopra, Khanna Book Publishing, 2018
2. Jeff Heaton, Deep Learning and Neural Networks, Heaton Research Inc, 2015

NATURAL LANGUAGE PROCESSING

Credits: 4 **Semester:III**
Subjectcode: MDS21307 **No. of lecture hours:60**
Prerequisites: Python, Machine Learning

Objectives:

- The goal is to become familiar with the concepts of the study of human language from a computational perspective.
- It covers syntactic, semantic and discourse processing models, emphasizing machine learning concepts.

Course Outcomes:

CO1: Understand various approaches on syntax and semantics in NLP.
CO2: Apply various methods to discourse, generation, dialogue and summarization using NLP.
CO3: Analyze various methodologies used in Machine Translation, machine learningtechniquesusedinNLPincludingunsupervisedmodelsandtoanalyze real timeapplications.

UNIT– I **12Hrs**

- Introduction to NLP, Background,and overview, 4
- NLP Applications -NLP hard Ambiguity- Algorithms And Models, 4
- KnowledgeBottlenecksinNLP-IntroductiontoNLTK,Case Study 4

UNIT– II **12Hrs**

Parsing and Syntax

- WordLevelAnalysis:RegularExpressions,TextNormalization, 4
- Edit Distance, Parsing and Syntax- Spelling, Error Detection, and correction-WordsandWordclasses-Part-ofSpeechTagging. 4
- Naive Bayes and Sentiment Classification:Case Study 4

UNIT– III **12Hrs**

Smoothed Estimation and Language Modelling

- N-gram LanguageModels: N-Grams, 3
- EvaluatingLanguageModels-The Language Modelling Problem 3

Semantic Analysis and Discourse Processing

- Semantic Analysis: Meaning Representation,LexicalSemantics, 3
- ambiguity, WordSenseDisambiguation. 3

UNIT– IV **12Hrs**

Natural language Generation and Machine Translation

- Natural Language Generation: Architecture of NLGSystems,Applications 3
- Machine Translation: Problems in Machine Translation- Machine Translation Approaches- Evaluation of MachineTranslationsystems. 3
- **Case study**: Characteristics ofIndianLanguages 3

UNIT– V **12Hrs**

Information Retrieval and Lexical Resources

- Information Retrieval: Design features of InformationRetrievalSystems 2
- Classical, Non- classical, Alternative Models ofInformationRetrieval, 2
- valuation Lexical Resources: Word Embeddings -Word2vec-Glove. 3
- **Recommended Systems, Long short-term memory (LSTM),** 3
- **Linear Discriminant Analysis(LDA)** 2

ESSENTIAL READING

1. SpeechandLanguageProcessing,DanielJurafskyandJamesH.,2ndEdition,Martin PrenticeHall,2013.
2. Foundations of Statistical NaturalLanguage Processing. Cambridge, MA: MIT Press, 1999.

SUGGESTED READING

1. Foundations of Computational Linguistics: Human-computer Communication in Natural Language, Roland R. Hausser, Springer,2014.
2. StevenBird,EwanKleinandEdwardLoperNaturalLanguageProcessingwith Python, O'Reilly Media; 1 edition,2009.

INTERNET OF THINGS LAB
(LAB-VII)

Credits:2 **Semester:III**
SubjectCode:MDS21308 **No. of lecture hours:45**

Objective:

- To Acquire the skill of IoT practical concepts.

Outcomes:

- To Implement IoT based Practicals.

Lab Exercises:

1. Setting up the Hardware and Software environment and library installation
2. Interfacing output devices
    a. LED
    b. Relay
    c. DCmotor
3. Interfacing input devices
    a. Switch
    b. IRsensor
4. Interfacing 16x2LCD
    a. Displaying Message
    b. Scrolling messages onLCD
    c. LM 35
5. Working with analog read and analog write functions.
    a. Reading data from Potentiometer and displaying it onLCD
    b. Interfacing Servo Motor and controlling usingPotentiometer.
6. Interfacing Ultrasonic sensor for safety parking system and creating an excel dataset.
7. Interfacing DHT sensor and recording the values and creating an excel dataset.
8. Working on different API for communication – ThingSpeak or Ubidots for Desktop platform and BLYNK for mobileplatform
9. Recording data from DHT, IR and Ultrasonic sensor onto ThinkSpeak or Ubidots and BLYNK
10. Controlling output devices from cloud platforms.

DEEP LEARNING LAB
(LAB-VIII)

Credits:2 **Semester:III**
SubjectCode: MDS23309 **No. of lecture hours:45**

Objective:
- To Build The Foundation Of Deep Learning.
- To Understand How To Build The Neural Network.
- To enable students to develop successful machine learning concepts.

Outcomes:
- Implement deep neural networks to solve real world problems.
- Choose appropriate pre-trained model to solve real time problem.
- Interpret the results of two different deep learning models.

Lab Exercises:
1. Implement multilayer perceptron algorithm for MNIST Hand written Digit Classification.
2. Design a neural network for classifying movie reviews (Binary Classification) using IMDB dataset.
3. Design a neural Network for classifying news wires (Multi class classification) using Reuters dataset.
4. Design a neural network for predicting house prices using Boston Housing Price dataset.
5. Build a Convolution Neural Network for MNIST Hand written Digit Classification.
6. Build a Convolution Neural Network for simple image (dogs and Cats) Classification
7. Use a pre-trained convolution neural network (VGG16) for image classification.
8. Implement one hot encoding of words or characters.
9. Implement word embedding for IMDB dataset.
10. Implement a Recurrent Neural Network for IMDB movie review classification problem.

TEXT BOOKS: Reza Zadeh and BharathRamsundar, “Tensorflow for Deep Learning”, O‟Reilly publishers, 2018

REFERENCES: https://github.com/fchollet/deep-learning-with-python-notebooks

NATURAL LANGUAGE PROCESSING LAB
(LAB-IX)

Credits:2 **Semester:III**
SubjectCode:MDS21310 **No. of lecture hours:45**

Objective:

- ToAcquiretheskillofmethodologiesusedinMachineTranslation,machine learning techniques used in NLP including unsupervised models and to analyze real timeapplications.

Outcomes:

- Implementthemethodologiesusedin MachineTranslation,machinelearning techniquesusedinNLPincludingunsupervisedmodelsandtoanalyzereal timeapplications.

Lab Exercises:

1. Implement and demonstrate the tokenize text.
2. Implementanddemonstratethecountwordfrequencyandtoremovestopwords.
3. Implement and demonstrate the tokenize Non-EnglishLanguages
4. Implement and demonstrate the get synonyms fromWordNet.
5. Implement and demonstrate to get Antonyms fromWordNet.
6. Implement and demonstrate the non-English Words
7. Implement and demonstrate the lemmatizing words UsingWordNet.
8. Implement and demonstrate the different stemming and lemmatization words.
9. Implement and demonstrate the POS Tagging or Word Embeddings.
10. Case study-based program on Sentiment Analysis

YEAR-WISE AND SEMESTER-WISE DISTRIBUTION OF SUBJECTS
M.Sc. Data Science FOURTH SEMESTER
ACADEMIC YEAR 2023-24 OF 2022-2024 BATCH(CBCS)

Sl. No.	Subject Code	Title of the Subject	Hours /week	Marks			Credits
				Internal	External	Total	
1	MDS21401	**Major Project**	30	50	150	200	15
Total			**30**	**50**	**150**	**200**	**15**

PROJECT

Credits: 15 **Semester: IV**
Subject code: MDS21401 **No. of hours: 90**

Objective:

- The primary emphasis of the project work is to understand and gain the knowledge of the principles of data science and analytics practices.

Course Outcome:

- The student will be able to describeData Collection, Data Visualization, Labeling, Data Selection, Data Pre-processing, Data transformation, Model Training, Model evaluation, Model testing, Accuracy, and deployment.

Fourth Semester of the M.Sc. Data Science course is exclusively meant for project work. Project must be carried out by each student individually in a period of 15 weeks. Students should submit a synopsis at the end of 3rdSemester in consultation with the Project Guide. The synopsis should consist of Data Collection, Data Visualization, Labeling, Data Selection, Data Pre-processing, Data transformation, Model Training, Model evaluation, Model testing,Accuracy and deployment are required to show in the project work.

A committee consisting of the department faculty members of the respective college along with a guide will evaluate the project and award internal marks.

At the end of the semester the students are required to present themselves for a Viva-voce examination in which each student will be awarded with an external mark.