B. Tech with

MINOR

in

COMPUTER SCIENCE AND ENGINEERING – AIML

Academic Regulations, Course Structure and Syllabus

Effective from 2023-24 admitted batches

Offered by

Department of Computer Science and Engineering

KSRM College of Engineering (A) – Kadapa

(Approved by AICTE, Accredited by NAAC with A+ Grade and NBA and Affiliated to JNTUA, Anantapuramu)

ELIGIBILITY / REGISTRATION / AWARD OF MINOR

The primary objective of a minor degree is to provide students with a secondary area of study to broaden their knowledge, enhance their skill set, and potentially improve their career prospects. It allows students to explore interests beyond their major, potentially leading to a more well-rounded and competitive profile.

i) Minor degree is introduced by the respective departments offering B. Tech. programs and is applicable to all B.Tech (Regular and Lateral Entry) students admitted in Engineering & Technology.

ii) There is NO minimum CGPA requirement to register for Minor degree program. However, the student shouldn't have any course backlog **up to III Semester in the major degree.**

iii) A student shall earn **additional 18 credits for award of Minor degree** offered by the department other than parent department. This is in addition to 163 credits by a regular student and 123 Credits by a Lateral Entry student for the award of Major degree.

iv) A student is permitted to register for Minor in IV Semester after the results of III Semester are declared. Students shall register and pass in all the courses prescribed and being offered from V semester under the respective Minor degree.

v) Students have to attend classwork for courses under Minor degree beyond regular academic hours meant for major degree. Students can also undergo the courses under Minor through any proctored online platforms with the prior approval of the BoS Chairman and the HoD of the respective department offering Minor degree.

vi) The attendance for the registered courses under Minor and regular courses offered for Major degree in a Semester will be considered separately.

vii) A student shall have an aggregate of 75% attendance in all courses registered under Minor in that particular semester to become eligible for attending Semester-End examinations.

viii) The registration for the Minor will be cancelled, if the student is detained due to lack of attendance in Major.

ix) The registration for the Minor will be cancelled, if the student fails in any course of either Minor / Major in any semester from V to VIII Semester.

x) A student registered for Minor shall pass in all subjects that constitute the requirement for the Minor degree program. No class/division (i.e., second class, first class and distinction, etc.) will be awarded for Minor degree program.

xi) A separate grade sheet will be issued for the Minor degree courses semester-wise..

xii) If a student drops or is terminated from the Minor program, the additional credits so far earned cannot be converted into open or core electives; they will remain extra.

xiii) The Minor will be mentioned in the degree certificate as Bachelor of Technology in Civil Engineering with Minor in Mechanical Engineering.

xiv) There shall be a minimum enrolment of 20% OR 20 enrollments from the list of

eligible students to offer Minor program.

xv) There is no fee for registration of courses for Minor program offered.

xvi) A student can register for either Minor / Honor but not both.

xvii) Student shall submit an application for either Minor / Honor at least one week before the commencement of the V Semester.

MINOR PROGRAMS OFFERED

Offering Department	Title	Who can Register
Civil Engineering	Civil Engineering	B.Tech. EEE / ME / ECE / CSE / AIML / CSE(AIML) / CSE(DS)
Mechanical Engineering	3D Printing	B.Tech. CE / EEE / ECE / CSE / AIML / CSE(AIML) / CSE(DS)
	Industrial Engineering	B.Tech. CE / EEE / ECE / CSE / AIML / CSE(AIML) / CSE(DS)
Electrical and Electronics Engineering	Energy Systems	B.Tech. CE / ME / ECE / CSE / AIML / CSE(AIML) / CSE(DS)
Electronics and Communication Engineering	Embedded Systems	B.Tech. CE / EEE / ME / CSE / AIML / CSE(AIML) / CSE(DS)
Computer Science and Engineering	Computer Science and Engineering	B.Tech. CE / EEE / ME / ECE
	Data Science	B.Tech. CE / EEE / ME / ECE
	Computer Science and Engineering – Artificial Intelligence and Machine Learning	B.Tech. CE / EEE / ME / ECE

COURSE STRUCTURE

for

MINOR

in

COMPUTER SCIENCE AND ENGINEERING – AIML

S.No	COURSE CODE	COURSE TITLE	Semester Offered	L	T	P	IM	EM	CR
1	2333571M	Computer Vision and Image Processing	V	3	0	0	30	70	3
2	2333572M	Data Visualization	V	3	0	0	30	70	3
3	2333573M	Computer Vision and Image processing Lab	V	0	0	3	30	70	1.5
4	2333671M	Big Data Technologies	VI	3	0	0	30	70	3
5	2333672M	Soft Computing	VI	3	0	0	30	70	3
6	2333673M	Big Data Technologies Lab	VI	0	0	3	30	70	1.5
7	2333771M	Data Wrangling	VII	3	0	0	30	70	3
			Total	**15**	**0**	**6**	**210**	**490**	**18**

2333571M	MINOR IN COMPUTER SCIENCE AND ENGINEERING – AIML COMPUTER VISION AND IMAGE PROCESSING (CE,ME,ECE,EEE)	L	T	P	C
		3	0	0	3

Pre-Requisites: Introduction to Programming, Python Programming, Probability and Statistics

Course Outcomes:

Upon successful completion of the course, students will be able to:

CO1. Describe image formation, representation, and apply basic image processing and frequency domain techniques for image enhancement and restoration.
CO2. Apply edge detection, segmentation, morphological, and texture analysis techniques for extracting features from images.
CO3. Analyze 3D vision and motion using techniques like stereo vision, optical flow, and camera calibration for scene understanding and depth estimation.
CO4. Evaluate object recognition approaches and machine learning models including traditional and deep learning techniques used in computer vision.
CO5. Implement advanced computer vision applications such as image compression, face recognition, and medical image analysis using case studies.

SYLLABUS:

UNIT - I: INTRODUCTION TO COMPUTER VISION AND IMAGE PROCESSING (09 Periods)

Overview of Computer Vision and Image Processing: Definitions and scope, Historical development and applications, Image Formation and Representation: Image acquisition methods, Sampling and quantization, Color spaces and models, Fundamentals of Image Processing: Point operations (brightness and contrast adjustments), Histogram processing, Spatial filtering techniques Fourier Transform and Frequency Domain Processing: Discrete Fourier Transform (DFT), Filtering in the frequency domain, Image restoration concept.

UNIT - II: IMAGE ANALYSIS TECHNIQUES (09 Periods)

Edge Detection and Feature Extraction: Gradient operators (Sobel, Prewitt), Canny edge detector, Corner and interest point detection, Image Segmentation: Thresholding methods, Region-based segmentation, Clustering techniques (K-means, Mean-Shift), Morphological Image Processing: Erosion and dilation, Opening and closing operations, Applications in shape analysis, Texture Analysis, Statistical methods (co-occurrence matrices), Transform-based methods (Gabor filters), Applications in pattern recognition.

UNIT - III: 3D VISION AND MOTION ANALYSIS (09 Periods)

Stereo Vision: Epipolar geometry, Disparity mapping, Depth estimation techniques, Structure from Motion (SfM): Feature tracking across frames, 3D reconstruction from

motion, Applications in scene understanding, Optical Flow and Motion Analysis: Lucas-Kanade method, Horn-Schunck method, Motion segmentation, Camera Calibration and 3D Reconstruction: Intrinsic and extrinsic parameters, Calibration techniques, 3D point cloud generation

UNIT - IV: OBJECT RECOGNITION AND MACHINE LEARNING IN VISION (10 Periods)

Feature Descriptors and Matching: Scale-Invariant Feature Transform (SIFT), Speeded-Up Robust Features (SURF), Feature matching algorithms, Object Detection and Recognition: Template matching, Deformable part models, Convolutional Neural Networks (CNNs), Introduction to Machine Learning for Vision: Supervised and unsupervised learning, Support Vector Machines (SVMs), Decision trees and random forests, Deep Learning Architectures: Autoencoders, Recurrent Neural Networks (RNNs), Generative Adversarial Networks (GANs)

UNIT - V: APPLICATIONS AND ADVANCED TOPICS (08 Periods)

Image Compression: Lossy and lossless compression techniques, Standards (e.g., JPEG, PNG), Morphological Image Processing: Dilation, erosion, opening, and closing operations., Applications in shape analysis, Case Studies: Face recognition systems., Automated visual inspection, medical image analysis.

Total Periods: 45

TEXTBOOKS:

T1. Digital Image Processing, R. C. Gonzalez, R. E. Woods, Pearson Prentice Hall, Stony Brook University, 3rd ed.,2008.
T2. Computer Vision: Algorithms and Applications, Richard Szeliski, Springer, 2010.

REFERENCE BOOKS:

R1 Computer Vision: A Modern Approach, David A. Forsyth, Jean Ponce, Prentice Hall, 2002.
R2 Computer Vision – Linda G. Shapiro, George C. Stockman, Prentice Hall, 2001.

ONLINE LEARNING RESOURCES:

1 Coursera: Introduction to Computer Vision and Image Processing. LinkCoursera
2 Stanford University: CS231n: Deep Learning for Computer Vision. Linkcs231n.stanford.edu
3 MIT OpenCourseWare: Introduction to Computer Vision. Link

2333572M	MINOR IN COMPUTER SCIENCE AND ENGINEERING – AIML DATA VISUALIZATION (CE,ME,ECE,EEE)	L	T	P	C
		3	0	0	3

Pre-Requisites: Introduction to Programming, Python Programming

Course Outcomes:

After completion of the course, students will be able to:

CO1. Interpret different types of data and recognize the appropriate visualization methods.
CO2. Design effective and interactive data visualizations using various tools.
CO3. Apply visual encoding and perceptual principles in presenting complex data.
CO4. Analyze and visualize real-world data sets using Python libraries and dashboards.
CO5. Create visual stories and dashboards for effective communication of insights.

SYLLABUS:

UNIT - I: INTRODUCTION TO DATA VISUALIZATION & PERCEPTION (09 Periods)

Introduction to Data Visualization, Importance and Scope of Data Visualization, Data Types and Sources, Visual Perception: Pre-attentive Processing, Gestalt Principles, Data-Ink Ratio, Data Density, Lie Factor, Visualization Process and Design Principles, Tools Overview: Tableau, Power BI, Python Libraries

UNIT - II: VISUALIZATION TECHNIQUES FOR CATEGORICAL & QUANTITATIVE DATA (09 Periods)

Charts for Categorical Data: Bar Charts, Pie Charts, Column Charts, Charts for Quantitative Data: Histograms, Line Charts, Boxplots, Scatter Plots, Bubble Charts, Heatmaps, Choosing the Right Chart Type, Best Practices in Labeling, Coloring, and Scaling.

UNIT - III: MULTIDIMENSIONAL, TEMPORAL AND HIERARCHICAL DATA VISUALIZATION (09 Periods)

Visualizing Multivariate Data: Parallel Coordinates, Radar Charts, Time-Series Visualization: Time Plots, Animation over Time, Geographic Data Visualization: Maps, Choropleths, Hierarchical Data: Treemaps, Sunburst Charts, Network and Graph Visualization.

UNIT - IV: DATA VISUALIZATION USING PYTHON AND DASHBOARDS (09 Periods)

Introduction to Matplotlib, Seaborn, and Plotly, Creating Static and Interactive Charts, Pandas Visualization Capabilities, Dashboards with Dash, Streamlit, Power BI, Case Studies: Real-world Dataset Visualization.

UNIT - V: STORYTELLING WITH DATA AND ETHICAL VISUALIZATION

(09 Periods)

Storytelling and Narrative Techniques in Visualization, Dashboards and Reporting, Misleading Visualizations and Bias, Ethical Principles in Data Visualization, Final Project: Create a Storytelling Dashboard with Real Data.

Total Periods :45

TEXTBOOKS:

T1. Visualization Analysis and Design – Tamara Munzner – CRC Press, 2014.
T2. Data Points: Visualization That Means Something – Nathan Yau – Wiley, 2013.

Reference Books:

R1 The Truthful Art: Data, Charts, and Maps for Communication – Alberto Cairo – New Riders, 2016.
R2 Storytelling with Data: A Data Visualization Guide for Business Professionals – Cole Nussbaumer Knaflic – Wiley, 2015.
R3 Fundamentals of Data Visualization – Claus O. Wilke – O'Reilly, 2019.
R4 Hands-On Data Visualization with Bokeh – Rohan Chopra – Packt Publishing, 2019.

ONLINE LEARNING RESOURCES:

1 NPTEL: Data Visualization - IIT Madras
2 Coursera: Data Visualization with Python by IBM

2333573M	MINOR IN COMPUTER SCIENCE AND ENGINEERING – AIML COMPUTER VISION AND IMAGE PROCESSING LAB (CE,ME,ECE,EEE)	L	T	P	C
		0	0	3	1.5

Pre-Requisites: Python Programming

Course Outcomes:

After successful completion of this lab, students will be able to:

CO1. Apply computer vision techniques to solve real-time image processing problems.
CO2. Train and evaluate machine learning models for classification and regression tasks.
CO3. Design and test feature extraction techniques from images.
CO4. Use OpenCV, Scikit-learn, TensorFlow/PyTorch for practical implementations.
CO5. Integrate vision-based features with ML algorithms for end-to-end solutions.

LIST OF EXPERIMENTS

1. Image preprocessing techniques: resizing, filtering, thresholding using OpenCV
2. Edge detection using Sobel, Canny, and Laplacian operators
3. Object detection using contour detection and bounding boxes
4. Feature extraction using HOG, SIFT, and ORB
5. Implement face detection using Haar cascades or DNN models
6. Train a machine learning model (SVM / KNN) for image classification
7. Build and evaluate a decision tree classifier using scikit-learn
8. Implement a logistic regression model for binary classification on numerical dataset
9. Apply PCA for feature reduction and visualization
10. Design a simple neural network using TensorFlow/Keras for image classification
11. Train and evaluate a CNN model for digit recognition using MNIST dataset
12. Real-time emotion recognition using webcam input and pre-trained model integration

TEXTBOOKS:

T1. Computer Vision: Models, Learning, and Inference, Simon J. D. Prince, Cambridge University Press.
T2. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, Aurélien Géron, O'Reilly, 2nd Edition.
T3. Computer Vision: Algorithms and Applications, Richard Szeliski, Springer.

REFERENCE BOOKS:

R1 Practical Python and OpenCV, Adrian Rosebrock, PyImageSearch.
R2 Deep Learning, Ian Goodfellow, Yoshua Bengio, Aaron Courville, MIT Press.
R3 Pattern Recognition and Machine Learning, C. M. Bishop, Springe.

ONLINE LEARNING RESOURCES:

1 https://opencv.org
2 https://www.tensorflow.org/tutorials
3 https://www.kaggle.com/learn/intro-to-machine-learning
4 https://www.pyimagesearch.com
5 NPTEL Course on Deep Learning

2333671M	MINOR IN COMPUTER SCIENCE AND ENGINEERING – AIML BIG DATA TECHNOLOGIES (CE,ME,ECE,EEE)	L	T	P	C
		3	0	0	3

Pre-Requisites: Database Management Systems

Course Outcomes:

CO1. Analyze the Bigdata characteristics
CO2. Make use of HDFS interfaces to read and write files
CO3. Analyze the data with MapReduce classes
CO4. Build the development environment of Hadoop to run the job on local job runner and on a cluster
CO5. Summarize the database applications of Hadoop

SYLLABUS:

UNIT - I: INTRODUCTION TO BIG DATA AND HADOOP ECOSYSTEM (09 Periods)

Introduction to Big Data, why is Big Data, Why Big Data is important, Meet Hadoop, Data, Data Storage and Analysis, Comparison with other systems, Grid Computing, A brief history of Hadoop, Apache Hadoop and the Hadoop Ecosystem, Linux refresher; VMWare Installation of Hadoop.

UNIT - II: HADOOP DISTRIBUTED FILE SYSTEM (HDFS) ARCHITECTURE AND OPERATIONS (09 Periods)

The Design of HDFS, HDFS Concepts, Command Line interface to HDFS, Hadoop File Systems, Interfaces, Java Interface to Hadoop, Anatomy of a file read, Anatomy of a file write, Replica placement and Coherency Model, Parallel copying with distcp, Keeping an HDFS cluster balanced.

UNIT - III: DATA ANALYSIS USING HADOOP AND MAPREDUCE PROGRAMMING (09 Periods)

Introduction, Analyzing data with unix tools, Analyzing data with Hadoop, Java MapReduce classes(new API), Data flow, combiner functions, Running a distributed MapReduce job, Configuration API, Setting up the developing environment, Managing configuration, Writing a unit test with MRUnit, Running a job in local job runner, Running on a cluster, Launching a job, The MapReduce WebUI.

UNIT - IV: ADVANCED MAPREDUCE PROGRAMMING AND PERFORMANCE TUNING (09 Periods)

Classic MapReduce, Job submission, Job initialization, Task Assignment,

Taskexecution, Progress and status updates, Job Completion, Shuffle and sort on Map and Reducer side, Configuration tuning, Map Reduce types, Input formats, Sorting, Map side and Reduce side joins.

UNIT - V: HIVE AND HBASE FOR BIG DATA PROCESSING (09 Periods)

The Hive Shell, Hive services, Hive clients, The meta store, comparison with traditional databases, Hive QI, Hbasics, Concepts, implementation, Java and Map reduce clients, Loading Data, Web queries.

Total Periods:45

TEXT BOOKS:

T1. Tom White, Hadoop, "The Definitive Guide" , 3rd Edition, O'Reilly Publications, 2012

T2. Dirk deRoos, Chris Eaton, George Lapis, Paul Zikopoulos, Tom Deutsch, Understanding Big Data Analytics for Enterprise class Hadoop and Streaming Data", 1st Edition, TMH, 2012.

REFERENCE BOOKS:

R1 Hadoop: The Definitive Guide by Tom White, O'Reilly Media

R2 Hadoop in Action by Chuck Lam, Manning Publications

R3 Big Data Analytics with R and Hadoop by Vignesh Prajapati, Packt Publishing

R4 Hadoop MapReduce Cookbook by Srinath Perera & Thilina Gunarathne, Packt Publishing

2333672M	MINOR IN COMPUTER SCIENCE AND ENGINEERING – AIML SOFT COMPUTING (CE,ME,ECE,EEE)	L	T	P	C
		3	0	0	3

Pre-Requisites: Machine Learning

Course Outcomes:

After completion of the course, students will be able to:

CO1. Understand the components and applications of soft computing.
CO2. Apply fuzzy logic concepts to real-world problems.
CO3. Build and train various neural network models.
CO4. Implement genetic algorithms for problem-solving and optimization.
CO5. Design hybrid systems using soft computing techniques.

SYLLABUS:

UNIT - I: INTRODUCTION TO SOFT COMPUTING AND FUZZY LOGIC (09 Periods)

Introduction to Soft Computing: Definition, Components, Differences with Hard Computing, Applications of Soft Computing, Fuzzy Logic: Crisp Sets vs Fuzzy Sets, Membership Functions, Fuzzy Set Operations, Fuzzy Rules and Fuzzy Reasoning, Fuzzy Inference Systems: Mamdani and Sugeno Models, Defuzzification Techniques.

UNIT - II: ARTIFICIAL NEURAL NETWORKS – I (09 Periods)

Introduction to Neural Networks: Biological Neurons vs Artificial Neurons, Architecture of Neural Networks: Feedforward, Feedback, Learning Rules: Hebbian, Delta, Perceptron Learning Rule, Single Layer Perceptron and its Limitations, Multi-Layer Perceptron: Backpropagation Algorithm, Applications of Neural Networks.

UNIT - III: ARTIFICIAL NEURAL NETWORKS – II (09 Periods)

Hopfield Networks and Associative Memories, Radial Basis Function Networks, Self-Organizing Maps (SOM), Recurrent Neural Networks (RNNs) – Basic Concepts, Convolutional Neural Networks (CNNs) – Overview and Applications, Practical Use Cases in Image and Pattern Recognition.

UNIT - IV: Genetic Algorithms and Optimization (09 Periods)

Introduction to Genetic Algorithms, GA Operators: Selection, Crossover, Mutation, Fitness Function and Evaluation, Schema Theorem, Elitism, Applications in Function Optimization, Scheduling, and Robotics, Introduction to Particle Swarm Optimization (PSO).

UNIT - V: HYBRID SYSTEMS AND ADVANCED TOPICS (09 Periods)

Hybrid Systems: Neuro-Fuzzy Systems, Fuzzy-GA, GA-ANN, ANFIS: Architecture and Learning, Case Studies on Hybrid Systems, Introduction to Deep Learning in Soft Computing, Real-World Applications: Forecasting, Control Systems, Medical Diagnosis, Image Processing.

Total Periods:45

TEXTBOOKS:

T1. Principles of Soft Computing, S. N. Sivanandam, S. N. Deepa, Wiley India, 3rd Edition
T2. Fuzzy Logic with Engineering Applications, Timothy J. Ross, Wiley, 4th Edition
T3. Neural Networks, Fuzzy Logic and Genetic Algorithm: Synthesis and Applications, S. Rajasekaran, G. A

REFERENCE BOOKS:

R1 Fundamentals of Neural Networks: Architectures, Algorithms and Applications, Laurene Fausett, Pearson
R2 Genetic Algorithms in Search, Optimization and Machine Learning, David E. Goldberg, Pearson
R3 Neural Networks and Learning Machines, Simon Haykin, Pearson, 3rd Edition
R4 Neural Networks and Fuzzy Systems, Bart Kosko, Prentice Hall

ONLINE LEARNING RESOURCES:

1 NPTEL – Soft Computing by Prof. S. Sengupta (IIT Kharagpur)
2 Coursera – Neural Networks and Deep Learning (Andrew Ng)

2333771M	MINOR IN COMPUTER SCIENCE AND ENGINEERING – AIML COMPUTER VISION AND IMAGE PROCESSING LAB (CE,ME,ECE,EEE)	L	T	P	C
		0	0	3	1.5

Pre-Requisites: Python Programming

Course Outcomes:

After successful completion of this lab, students will be able to:

CO1. Apply computer vision techniques to solve real-time image processing problems.
CO2. Train and evaluate machine learning models for classification and regression tasks.
CO3. Design and test feature extraction techniques from images.
CO4. Use OpenCV, Scikit-learn, TensorFlow/PyTorch for practical implementations.
CO5. Integrate vision-based features with ML algorithms for end-to-end solutions.

LIST OF EXPERIMENTS

1 Image preprocessing techniques: resizing, filtering, thresholding using OpenCV
2 Edge detection using Sobel, Canny, and Laplacian operators
3 Object detection using contour detection and bounding boxes
4 Feature extraction using HOG, SIFT, and ORB
5 Implement face detection using Haar cascades or DNN models
6 Train a machine learning model (SVM / KNN) for image classification
7 Build and evaluate a decision tree classifier using scikit-learn
8 Implement a logistic regression model for binary classification on numerical dataset
9 Apply PCA for feature reduction and visualization
10 Design a simple neural network using TensorFlow/Keras for image classification
11 Train and evaluate a CNN model for digit recognition using MNIST dataset
12 Real-time emotion recognition using webcam input and pre-trained model integration

TEXTBOOKS:

T1. Computer Vision: Models, Learning, and Inference, Simon J. D. Prince, Cambridge University Press.
T2. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, Aurélien Géron, O'Reilly, 2nd Edition.
T3. Computer Vision: Algorithms and Applications, Richard Szeliski, Springer.

REFERENCE BOOKS:

R1 Practical Python and OpenCV, Adrian Rosebrock, PyImageSearch.
R2 Deep Learning, Ian Goodfellow, Yoshua Bengio, Aaron Courville, MIT Press.
R3 Pattern Recognition and Machine Learning, C. M. Bishop, Springe.

ONLINE LEARNING RESOURCES:

1 https://opencv.org
2 https://www.tensorflow.org/tutorials
3 https://www.kaggle.com/learn/intro-to-machine-learning
4 https://www.pyimagesearch.com
5 NPTEL Course on Deep Learning

2333771M	MINOR IN COMPUTER SCIENCE AND ENGINEERING – AIML DATA WRANGLING (CE,ME,ECE,EEE)	L	T	P	C
		3	0	0	3

Pre-Requisites: Python Programming, Data Visualization

Course Outcomes:
After successful completion of this course, students will be able to:

CO1. Describe and apply foundational data wrangling and acquisition techniques using Python and SQL.
CO2. Clean and preprocess data by handling missing values, inconsistencies, outliers, and standardization.
CO3. Transform and engineer features from raw datasets using encoding, parsing, reshaping, and grouping.
CO4. Merge and join multiple datasets, resolve schema mismatches, and automate preprocessing workflows.
CO5. Utilize advanced libraries/tools and apply best practices to build reproducible, clean datasets for ML.

SYLLABUS:

UNIT - I: INTRODUCTION TO DATA WRANGLING AND DATA ACQUISITION (09 Periods)

Introduction to Data Wrangling: Importance and Use Cases , Types of Data: Structured, Semi- Structured, Unstructured, Data Acquisition Techniques: APIs, Web Scraping, Reading Data from CSV, Excel, JSON, XML, Using Python libraries: pandas, requests, BeautifulSoup, Working with Databases using SQLAlchemy and pandas, Loading Large Datasets and Chunking, Exploratory Analysis Before Cleaning.

UNIT - II: HANDLING MISSING, NOISY, AND INCONSISTENT DATA (09 Periods)

Identifying and Understanding Missing Data, Techniques for Imputing Missing Values, Handling Inconsistent Data: Dates, Texts, Units, Removing Duplicates and Irrelevant Data, Detecting and Treating Outliers, Normalization and Standardization Techniques, Regular Expressions for Text Cleaning, Visualizing Missing/Outlier Data.

UNIT - III: DATA TRANSFORMATION AND FEATURE ENGINEERING (09 Periods)

Data Type Conversion and Parsing, Feature Extraction from Text, Dates, and Strings, One-Hot Encoding, Label Encoding, Binning and Discretization, Data Aggregation and Grouping, Pivoting, Melting, and Reshaping Data, Handling Imbalanced Data, Creating Derived Features and Feature Selection.

UNIT - IV: DATA INTEGRATION, JOINING, AND WORKFLOWS

(09 Periods)

Merging and Joining Datasets (Inner, Outer, Left, Right), Concatenation and Appending DataFrames, Data Consistency and Referential Integrity, Resolving Schema Mismatches, Designing Reusable Data Wrangling Functions, Automating Workflows with Functions and Pipelines, Data Lineage and Documentation, Case Study: End-to-End Data Wrangling Pipeline.

UNIT - V: TOOLS, LIBRARIES, AND CASE STUDIES IN DATA WRANGLING
(09 Periods)

Pandas and NumPy Advanced Techniques, Pyjanitor, Dask, and Polars for Efficient Wrangling, Using OpenRefine for Data Cleaning, SQL vs NoSQL in Data Wrangling, Real-world Wrangling Case Studies (Finance, Healthcare, Retail), Best Practices and Common Pitfalls in Data Wrangling, Reproducibility and Versioning in Data Pipelines, Final Capstone: Build and Evaluate a Clean Dataset for ML.

Total Periods:45

TEXTBOOKS:

T1. Data Wrangling with pandas, M. Heydt, O'Reilly Media.
T2. R for Data Science (Data Wrangling Chapters), Hadley Wickham, O'Reilly.
T3. Python Data Science Handbook, J. VanderPlas, O'Reilly Media.

REFERENCE BOOKS:

R1 Python for Data Analysis, Wes McKinney, O'Reilly.
R2 Doing Data Science, Cathy O'Neil, Rachel Schutt, O'Reilly.
R3 Cleaning Data for Effective Data Science, David Mertz, Packt.

ONLINE LEARNING RESOURCES:

1. Data Wrangling with pandas (Datacamp): https://www.datacamp.com/courses/data- manipulation-with-pandas
2. Coursera: Data Wrangling, Analysis and AB Testing with SQL – https://www.coursera.org/learn/data-wrangling-analysis-abtesting
3. edX: Data Wrangling with R – https://online.rice.edu/courses/data-wrangling-r
4. Real Python Tutorials on pandas: https://realpython.com/learning-paths/pandas/
5. Kaggle Notebooks (Data Cleaning & Wrangling): https://www.kaggle.com/learn/pandas