MCA473A – ADVANCED DATA ANALYTICS

Total Teaching Hours for Semester: 75 **Max Marks: 150**

Credits: 5

Course Objectives

This course introduces the fundamentals of Text and Image data processing. This course is designed to explore various methods and concepts in social media data and models, representation, various operations, transformation, restoration, and segmentation in Image processing. Also helps to learn how to apply a wide range of classification and clustering algorithms.

Course Outcomes

Upon successful completion of the course, the student will be able to

CO1: Understand the fundamentals of Text and Image Data.

CO2: Apply principles and techniques to social media and image data

CO3: Analyze and implement pre-processing algorithms

CO4: Implement classification and Clustering techniques for real-time text and image data.

Unit-1

INTRODUCTION **Teaching Hours: 15**

Overview of data analytics, Introduction to advanced data analytics techniques,Need for Advanced Data Analytics, Statistical Methods for Data Analysis:Probability distributions, Hypothesis testing, Regression analysis, Correlation and Covariance, Sampling and Estimation, Bayesian Statistics,Time series analysis, Complexities of modern datasets, Recent Technologies and Frameworks for Data Analytics, Role of data analytics in Text,Social Media and Image.

Lab Exercises:

1. Implementation of Correlation and Regression
2. Implementation of time series analysis

Unit-2 **Teaching Hours: 15**

TEXT ANALYTICS

Text Representation- tokenization, stemming, stop words, TF-IDF, NER, N-gram modeling. **Mining Textual Data:** Text Clustering, Text Classification,

Lab Exercises:

0. Implementation of tokenization, stemming, stop words
0. Implementation of TF-IDF, NER and N-gram

Unit-3 **Teaching Hours: 15**

SOCIAL MEDIA ANALYTICS

Essentials of Social Graphs, Social Networks, Models, Information Diffusion in Social Media. **Analyzing social media:** Behavioral Analytics, Influence, and Homophily, Recommendation in social media.

Lab Exercises:

0. Implementation of user Behavioral Analysis on any social media
0. Implementation of Clustering and Classification in text document/social media data

Unit-4 **Teaching Hours: 15**

IMAGE ANALYTICS

Digital Image Representation - Elements of digital image processing-Digital Image Properties-Histograms, Entropy -Relationships between pixelsConnectivity, Distance Measures between pixels -Various image formats – bmp, jpeg, tiff, png, gif. Noise in Images – Sources, types, Image Restoration, Image Filtering-Inverse filtering, Wiener Filtering - Segmentation

Lab Exercises:

0. Digitization and Implementation of Histogram Equalization
0. Implementation of metrics for Noise measures in Image quality

Unit-5 **Teaching Hours: 15**

CASE STUDIES:

Healthcare: Image Analytics/Video Analytics for health image/video, Predictive Analytics for Patient Admissions - Location-based Case study using GIS - Education: Predictive Analytics for Student Success - E-commerce: Personalized Recommendations - Manufacturing: Predictive Maintenance for Equipment

Lab Exercises:

0. Implementation of Image-filtering techniques
0. Implementation of Image classification and Image Clustering

Text Books and Reference Books

[1] John Atkinson-Abutridy, Text Analytics: An Introduction to the Science and Applications of Unstructured Information Analysis, CRC Press, 2022

[2] Rafael C. Gonzalez and Richard E. Woods, Digital Image Processing, Third Ed., Prentice-Hall,2020

[3] Sonka, Fitzpatrick, Medical Image Processing and Analysis, 2020.

Essential Reading / Recommended Reading

[1] William K. Pratt, Digital Image Processing, John Wiley, 4th Edition, 2020. 2. Anil K. Jain

[2] Fundamentals of Digital Image Processing, Prentice Hall of India,2020

Web Resources:

1. Image Analytics | NIST

CO – PO Mapping

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	2	2			2						1	
CO2	2	1	1		1	1					1	1

CO3	3	1	2	2	1						2	2
CO4	1	2	2	2	2	1					2	2