|  |  |
| --- | --- |
| **SESSION** | **FEB MARCH 2025** |
| **PROGRAM** | **MASTER OF BUSINESS ADMINISTRATION (MBA)** |
| **SEMESTER** | **IV** |
| **COURSE CODE & NAME** | **DADS402 UNSTRUCTURED DATA ANALYSIS** |
|  |  |
|  |  |

**Assignment Set – 1**

**Q1. (a) Summarize the various methods to store unstructured data.**

**(b) Interpret the difference between Text and Big data. 5+5**

**Ans 1.**

**a. Storage Methods for Unstructured Data**

Unstructured data refers to information that lacks a predefined data model or organizational structure. This includes text documents, images, videos, and social media content. Several storage methods are used to efficiently store and manage unstructured data. The first is the use of data lakes, which allow raw data to be stored in native format and are commonly hosted on cloud platforms like Amazon S3 and Azure Data Lake. Data lakes are highly scalable and support real-time analytics. Second, NoSQL databases such as MongoDB and

Its Half solved only

Buy Complete assignment from us

**Price – 190/ assignment**

**MUJ Manipal University Complete SolvedAssignments MARCH 2025**

buy cheap assignment help online from us easily

we are here to help you with the best and cheap help

**Contact No – 8791514139 (WhatsApp)**

**OR**

**Mail us-** bestassignment247@gmail.com

**Our website -** [www.assignmentsupport.in](http://www.assignmentsupport.in)

**Q2. (a) Illustrate the Naive Bayes classifier and how it works in text classification.**

**(b) Articulate a Machine Learning approach in sentiment analysis. Give a suitable example. 5+5**

**Ans 2.**

**a. Naive Bayes Classifier in Text Classification**

Naive Bayes is a supervised learning algorithm based on Bayes’ Theorem, often used for classification tasks in natural language processing. It is especially effective in text classification due to its simplicity, efficiency, and relatively high accuracy, even with limited training data. The core idea of Naive Bayes is to calculate the probability of a class based on the features of input data, assuming that all features are conditionally independent given the

**Q3. a. Describe Latent Dirichlet Allocation (LDA).**

**b. Discuss how NoSQL databases are different from relational databases. 5+5**

**Ans 3.**

**a. Latent Dirichlet Allocation (LDA)**

Latent Dirichlet Allocation (LDA) is a generative probabilistic model commonly used in topic modeling, which aims to uncover hidden thematic structures within a large collection of documents. The fundamental idea behind LDA is that documents are composed of multiple topics, and each topic is a distribution of words. By analyzing the patterns of word co-occurrence, LDA helps to identify latent topics that best describe a set of documents.

LDA assumes a two-level generative process. For each document, a topic distribution is first

**Assignment Set – 2**

**Q4. (a) Demonstrate how MongoDB ensures high availability and fault tolerance.**

**(b) Reframe Fast Fourier Transform (FFT). 5+5**

**Ans 4.**

**a. MongoDB and High Availability**

MongoDB is designed to deliver high availability and fault tolerance through its replica set architecture. A replica set in MongoDB is a group of mongod instances that maintain the same dataset. One of the nodes acts as the primary node, while others act as secondary nodes. The primary node receives all write operations, and the secondary nodes replicate the data from the primary.

This setup ensures high availability, as the system can automatically elect a new primary if

**Q5. (a) Discuss audio data preprocessing in machine learning.**

**(b) Connect how does histogram equalization work. 5+5**

**Ans 5.**

**a. Audio Data Preprocessing in Machine Learning**

Audio data preprocessing is a crucial step in preparing sound recordings for machine learning models. Raw audio signals contain noise, varying amplitudes, and redundant information that can hinder model performance. The goal of preprocessing is to convert these raw waveforms into structured features that algorithms can effectively learn from.

The first step involves resampling the audio to ensure all inputs have a uniform sample rate.

is critical for accurate image analysis and classification.

**Q6. (a) Extract the key components of a CNN for image classification.**

**(b) Conclude on some common techniques used for video classification.**

**Ans 6.**

**a. Components of a CNN for Image Classification**

Convolutional Neural Networks (CNNs) are a specialized class of deep learning models designed to process data with a grid-like topology, such as images. They are widely used in image classification tasks due to their ability to automatically extract hierarchical features from raw pixel data.

The input layer of a CNN receives the image, typically represented as a 2D or 3D array of pixel values. The first major component is the convolutional layer, which applies filters