**Unit -1(Feb – 2022, November—2023, )**

1. Elaborate different criteria on the basis of which data mining Techniques are classified
2. By Applying Min Max Normalization transform the 73000 with minimum and maximum values for income attribute 12000 and 98000 respectively to new range [0.0 , 1.0]
3. Elaborate the identification problem faced while integrating the data
4. With an apt example difference between supervised and unsupervised discretization
5. Define Concept hierarchy and its types ? Explain How it can be used in data reduction with example
6. How Sampling of data can improve Quality of data mining ? Difference Between Simple random Sampling without replacement and with replacement
7. How efficient and scalable frequent itemsets can be mined from given data set by apriori algorithm ? Explain the importance of prune steps in improving its efficiency

Unit – 2

1. State the importance of training and testing phase of any classification approach
2. Difference between active and lazy learners with two examples of each
3. How bagging and boosting methods can be used to improve classifiers performance
4. Elaborate how DBSCAN density based algorithm can be used for handling noise effectively

Unit – 3

1. Explain different types of digital data along with their data access methods and data management mechanism respectively of each type
2. Elaborate 7V’s responsible for making data handling and challenging task
3. What do you mean by serialization and deserialization and its importance in Hadoop Working environment
4. Elaborate the HDFS architecture and list down different daemons working in HDFS Cluster

Unit – 4

1. Detail Working of three major components Code Driver , Mapper and Reducer of Map Reduce framework
2. List various functions performed by Pig execution engine to successfully run the pig queries
3. Why partitions and buckets are created in Hive ? Explain with the help of example . Adding , renaming , and dropping a partition in hive
4. Elaborate how a request from a client is propagated in HBase architecture that is handled bu its different Components