G H Patel College of Engineering and Technology

Department Of Computer Engineering

**A.Y. 2018-19(ODD), Semester 7**

**Subject Name:** DATA MINING AND BUSINESS INTELLIGENCE

**Active Learning Assignment Topics**

**Note: Students should prepare Active Learning Assignment in group of two. (Each group has to pick unique topic from assigned topics)**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Enrolment No.** | **Topic** |
|  | 140110107061,22 150110107001,2 | Data Warehousing and Business Intelligence-What is Business Intelligence, BI and DW in today’s perspective, What is data warehousing, Defining Features, data marts, model evaluation, Estimating accuracy of classifier, Techniques to improve Classification accuracy |
|  | 150110107003,4,5,6 | BI and DW architectures and its types - Relation between BI and DW - OLAP (Online analytical processing) definitions - Difference between OLAP and OLTP. |
|  | 150110107007,8,9,10 | Defining schemas: Stars, snowflakes and fact constellations, Motivation for Data Mining - Data Mining-Definition and Functionalities. |
|  | 150110107011,12,13,14 | Why to pre-process data? - Data cleaning: Missing Values, Noisy Data - Data Integration and transformation. |
|  | 150110107015,16,17,18,19 | Data Reduction: Data cube aggregation, Dimensionality reduction - Data Compression - Numerosity Reduction |
|  | 130110107027,150110107020,  22,23,24 | What is concept description? - Data Generalization and summarization-based characterization - Attribute relevance - class comparisons, Association Rule Mining: Market basket analysis - basic concepts - |
|  | 150110107025,27,28,29 | Data mining for business Applications like Balanced Scorecard, Fraud Detection, Clickstream Mining, Market Segmentation, retail industry, telecommunications industry, banking & finance and CRM etc. |
|  | 150110107030,31,32,33 | Clustering, Spatial mining, web mining, text mining. |
|  | 150110107034,35,36,37,38,39 | Big Data: Introduction to big data: distributed file system – Big Data and its importance, Four Vs, Drivers for Big data, Big data analytics, Big data applications. Algorithms using map reduce, Matrix-Vector Multiplication by Map Reduce. Introduction to Hadoop architecture: Hadoop Architecture, Hadoop Storage: HDFS, Common Hadoop Shell commands. |
|  | 150110107040,41,42,43 | What is classification and prediction? – Issues regarding Classification and prediction: Classification methods: Decision tree, Bayesian Classification, Rule based. |
|  | 150110107044,45,46,47 | CART, Neural Network, Prediction methods: Linear and nonlinear regression, Logistic Regression Introduction of tools such as DB Miner /WEKA/DTREG DM Tools. |
|  | 150110107048,49,50,51,  130110107062 | Introduction to Big data Business Analytics - State of the practice in analytics role of data scientists Key roles for successful analytic project - Main phases of life cycle. |
|  | 150110107052,53,54,55,56,57 | Apriori algorithm - generating rules – Improved Apriori algorithm, FP growth algorithm, Associative Classification – Rule Mining. |
|  | 150110107058,59,60,61 | Data Mining Primitives - Languages and System Architectures: Task relevant data - Kind of Knowledge to be mined - Discretization and Concept Hierarchy. |
|  | 150110107062,63,64,  160113107001 | Classification of DM Systems - DM task primitives - Integration of a Data Mining system with a Database or a Data Warehouse, KDD Process. |
|  | 160113107002,3,4,5 | Dimensional analysis - What are cubes? Drill-down and roll-up - slice and dice or rotation, OLAP models - ROLAP versus MOLAP. |
|  | 160113107006,7,8,9,10,11 | Overview of the components - Metadata in the data warehouse - Need for data warehousing - Basic elements of data warehousing - trends in data warehousing, Issues in DM. |