Text

Description automatically generated

**Department of Computer Science and Engineering**

**18CS54 – Data Mining**

SEE Revision Topics:

**Unit 1:**

1. KDD Process - Diagram with explanation
2. Motivational Challenges
3. Data Mining Tasks and its applications
4. Privacy in Data Mining
5. Refer Lecture slides, MSE 1 QP
6. End of chapter exercises shared in GitHub

**Unit 2:**

1. Types of Data
2. Data Pre-processing
3. Measures of Similarity and Dissimilarity
4. Refer Lecture slides, MSE 1 QP
5. End of chapter exercises shared in GitHub

**Unit 3:**

1. Decision Tree Induction, (theory – Induction process, Gain Ratio & numericals – Information gain, Gini Index)
2. Rule Based Classifiers, (theory & numericals with likelihood, laplace and m-estimate measures) – Exercise problems discussed in lecture slides
3. Nearest Neighbor classifiers (numerical)
4. Refer Lecture slides, MSE 2 QP

**Unit 4:**

1. Apriori Algorithm- Frequent item set generation, Rule Generation (theory &numerical -FkxFk-1)
2. FP Growth Algorithm (theory & numerical)
3. Evaluation of Association Patterns (numerical: support, confidence, interest, IS & Correlation Coefficient)
4. Refer Lecture slides, MSE 2 & MSE 3 QP, Exercises discussed.

**Unit 5:**

1. Clustering Methods, comparison, specialty & limitations
2. K-Means (theory & numerical)
3. K-Medoid (theory)
4. Hierarchical Clustering:
   1. Agglomerative Clustering (theory & numerical)
   2. Divisive(theory)
5. BIRCH (theory)
6. DBSCAN (theory)
7. Refer Lecture slides, MSE3 QP & Exercises discussed

**Note:** For all the Units, please refer the lecture slides, MSE1-3 Question papers as well as the textbook contents with end of chapter exercises as per the prescribed syllabus to secure 100/100 in SEE exam.

All the best!