Lab Report

BSC.cSIT 7th Sem

Data Warehouse and Data Mining

Questions:

1. Explain the differences between an operational database and a data warehouse. Give at least one example of each.
2. List the main functionalities of a data mining system and explain one real-world application.
3. What is data cleaning? Describe two common techniques used to clean data before mining.
4. Define an “iceberg cube” and explain how it differs from a full data cube.
5. What are overfitting and underfitting in classification models? How does k-fold cross-validation help reduce these issues?
6. Compare partitioning clustering (like k-means) with hierarchical clustering. Give one example of each.
7. Using the dataset below, find all frequent itemsets with a minimum support of 50%. Frequent Itemset Mining(Apiriori Algorithm).

TransactionID	**Items**
1	Milk, Bread, Butter
2	Bread, Butter
3	Milk, Bread
4	Milk, Bread, Butter
5	Bread

8. Build a decision tree to predict whether a student passes based on Math and Science scores. Evaluate the accuracy using precision and recall.

StudentID	Math	Science	Passed
1	85	90	Yes
2	70	65	No
3	80	78	Yes
4	60	55	No
5	90	95	Yes

9. Perform k-means clustering on customer data and visualize the clusters. Use 2–3 clusters.

CustomerID	Annual Income (k$)	Spending Score (1-100)
1	15	39
2	16	81
3	17	6
4	18	77
5	19	40
6	20	76
7	21	6
8	22	94

10. Extract keywords and perform a simple term frequency analysis from the following documents.

DocumentID	Text
1	"Data mining is the process of discovering patterns in large data."
2	"Text mining helps in extracting useful information from documents."
3	"Classification and clustering are key tasks in data mining."
4	"K-means clustering groups similar data points together."
5	"Apriori algorithm finds frequent itemsets in transaction data."