Alexandria University  Course title: Data Mining

Data Mining project

- Problem Statement:

State the problem you aim to address through data mining.

- Dataset Selection:

Choose a dataset relevant to your problem statement. Consider datasets from various domains, such as healthcare, finance, marketing, or social media.

- Data Preprocessing:

Clean and preprocess the dataset to handle missing values, outliers, and inconsistencies. Perform data normalization or scaling as required

- Exploratory Data Analysis (EDA):

Conduct an exploratory analysis of the dataset to gain a better understanding of its structure and characteristics. Visualize the distributions, correlations, and statistical summaries of the variables. Identify any interesting patterns or outliers.

❖ **Apply 2 algorithms only**

➢ K-medoids Clustering:

Implement the k-medoids clustering algorithm on the dataset. Determine the optimal number of clusters using techniques. Assign each data point to a cluster based on its similarity to the medoid (representative) of the cluster.

➢ Hierarchical Clustering:

Utilize hierarchical clustering methods, such as agglomerative or divisive clustering, on the same preprocessed dataset. Construct a dendrogram to visualize the hierarchical relationships between data points. Determine the appropriate number of clusters based on the dendrogram's structure and the problem's requirements.

➢ Fuzzy Logic Clustering:

Implement fuzzy logic clustering on the preprocessed dataset. Convert numerical attributes into linguistic variables using fuzzy membership functions. Define appropriate fuzzy rules to infer the membership of each data point in different clusters based on their fuzzy values.

➢ Evaluation and Interpretation:

Evaluate the results of each clustering technique individually and collectively. Assess the quality of the clusters based on internal validation metrics and external criteria. Compare and contrast the

clustering outcomes from each technique to gain a comprehensive understanding of the dataset's underlying patterns and structures.

❖ The instructions that must be followed:

1. Your team has from 5 to 6 members.
2. Each team will make a report include name of member , id , and his role of project
3. Report must include Introduction Provide an overview of the project and its objectives ,Problem Statement Clearly state the problem you aim to address through data mining techniques ,screenshots of code and conclusion of project results
4. Deadline will be at 3-5-2024 , you will upload your report and code in attached form