Assignment 1: "Clustering Application with the Wisconsin Breast Cancer Dataset"

Purpose

- Implement clustering algorithms using data mining techniques for breast cancer diagnosis.
- This should include data preprocessing, model building, evaluation, and visualization steps.

Dataset to be Used

- Dataset: Wisconsin Breast Cancer Dataset
- Features: 30 numerical features (e.g., radius, texture, perimeter, area, smoothness)
- Classes: Malignant and Benign

Topics to be Performed:

1. Data Preparation
   - Class labels are extracted (since clustering is unsupervised learning)
2. Algorithms
   - K-Means
   - DBSCAN
3. Determining the Number of Clusters
   - Elbow method
   - Silhouette score
4. Evaluation
   - Comparison of cluster labels with actual classes
5. Visualization
   - Distribution of features
   - 2D scatter plot representation of cluster results

Files to be uploaded to the Sakai system:

- Jupyter Notebook: Python code and output
- Report: PDF format, method descriptions, and results evaluations