Ali Saleem (23L-2638)

Mini-Project: Human Activity Recognition Using WISDM Dataset

Overview of the Data

The dataset used for this project consists of accelerometer data collected from users performing different physical activities. It includes six columns: User_ID, Activity_Label, Timestamp, and three accelerometer readings (X, Y, Z). The dataset contains 19,142 entries and represents activities like walking, jogging, sitting, and standing. The goal was to classify these activities using various machine learning models.

Key Findings from Each Model

- **KNN (K-Nearest Neighbors):**
  KNN showed moderate accuracy (around 75%) in classifying the activities. While the training phase was fast since KNN is a lazy learner, the model can be slow during prediction, especially with larger datasets, as it calculates distances for each prediction.

- **K-Means/K-Medoids:**
  K-Means, used for clustering, was the fastest model to train but showed lower accuracy (around 70%) when used for classification. It struggled with overlapping activity patterns, making it less suitable for this type of supervised problem. Its strength lies in grouping data but not in precise classification.

- **SVM (Support Vector Machine):**
  SVM provided the highest accuracy, making it the best performer overall for activity classification. It handled high-dimensional data well, establishing clear decision boundaries between activities. However, SVM was the slowest to train due to its complex optimization process.

Considering the dataset's nature, SVM appears to be the best approach due to its high accuracy in classifying activities. While it’s slower to train, its ability to create distinct decision boundaries between different activities makes it ideal for this type of task. However, for scenarios where training time is critical, KNN could be a good alternative, though its accuracy is slightly lower K-Means, while quick to train, is better suited for clustering tasks rather than classification. Overall, SVM provides the most reliable performance for this dataset.