**MACHINE LEARNING & DATA MINING LAB (MLDM)**

**GROUP MEMBER : -**

SHIVANI TYAGI (17303011)

UTKARSHA (17303002)

**PROJECT TOPIC :** DATA MINING PROJECT ABOUT RECOGNIZING HUMAN ACTIVITY

**ABOUT PROJECT : -**

In this project, we implemented and evaluated classification algorithm to detect four crucial human physical activities (walking, cycling, sitting, and lying) using five triaxial accelerometers worn concurrently on different parts of the body (dominant hip, upper arm, ankle, thigh, and wrist). The accelerometer data were collected, cleaned, and pre-processed to extract features from 10 s window.

(The aim of project is to recognize human activity using multiple wearable accelerometer sensors placed at different body positions). These time and frequency domain features were used with Random Forest and k-Nearest Neighbour classifier to classify subject activities. The algorithms were evaluated based on Leave-One-Subject-Out (LOSO) and ten-fold cross-validation strategy using both accelerometer data as well as annotated activity labels from 33 participants in a lab.

Random Forest showed the best performance recognizing the activities with overall accuracy of 89 % for LOSO strategy for hip data. Combining data from both hip and ankle improved the overall accuracy by 3.5 %, and by 10% for lying activity, which had the lowest classification accuracy (80%) for hip data.

* Language: Python
* Tools: IPython, PyCharm
* Libraries: Scikit-Learn/SciPy, NumPy, Pandas, Matplotlib
* Algorithms : Random Forest, K-NN
* Evaluation Methodology: LOSO(Leave One Subject Out), K-Fold validation, Confusion Matrix

**Reference taken from :**  North-eastern University & GitHub

**MACHINE LEARNING & DATA MINING LAB (MLDM)**

**GROUP MEMBER : -**

TUHINA SHREE (17303019)

ABHA (17303018)

**PROJECT TOPIC :** TEXT MINING

**ABOUT PROJECT : -**

In this project, there are two major tasks: text data processing and text categorization.

In text data processing, we have done tokenization, stemming, normalization, etc. Also, vector space model and statistical language models are used to retrieve similar documents to query.

In text categorization, we build a text classification system which includes feature selection, classifiers (Naive Bayes and K Nearest Neighbour using brute force and random vectors), cross validation, and parameter tuning.

* Language: Python
* Tools: IPython, PyCharm
* Libraries: Scikit-Learn/SciPy, NumPy, Pandas, Matplotlib
* Algorithms : K-NN, Naïve Bayes , Cross Validation & Parameter Tuning
* Evaluation Methodology: Brute Force & Random Vectors, Confusion Matrix