**PROBLEM STATEMENT**

Processing of huge amounts of data and high-complexity computations are the two emerging major challenges faced by in the above six researches. In the above analysis we gained in-depth understanding of the activity classification system and data mining concepts, and provide recommendation based on observed a problem in finding an optimal solution for multiclass classification.

**OBJECTIVE:**

We need to explore One-Vs-Rest algorithm for multiclass classification. This algorithm involves training a single classifier per class, with the samples of that class as positive samples and all other samples as negative samples. This strategy requires the base classifiers to produce a real-valued confidence score for its decision, rather than just a class label; discrete class labels alone can lead to ambiguities, where multiple classes are predicted for a single sample**.** One-Vs-Rest algorithm takes a binary classifier as a parameter. We tried SVM, Logistic Regression, and LDA as the binary classifiers. Out of which LDA gave the best accuracy but not as good as Random Forest and kNN algorithm. So, we should focuss on RF and kNN to build a robust classifier.

**EVALUATION METHODOLOGY:**

LOSO(Leave One Subject Out), K-Fold validation, Confusion Matrix

**ALGORITHMS USED:**

The parameters of the estimator used to apply these methods are optimized by cross-validated grid-search over a parameter grid. Using these techniques we will apply the following parameters for RF and k-NN:

● Random Forest

○ Number of Decision Tree created : 50-200

○ Algorithm to be used : InfoGain(entropy)

● k-NearestNeighbor

○ Value of k varied from 9 - 11

○ Distance to be measured : euclidean

○ uniform weight for all the points

**ENVIRONMENT REQURIED:**

Python’s scikit-learn package will be used for performing all of the above mentioned tasks. This package provides standard machine learning libraries that we can modify as per our need and use it for our dataset. In addition to this, we will be using numpy, pandas for data processing and matplotlib for visualization purpose.

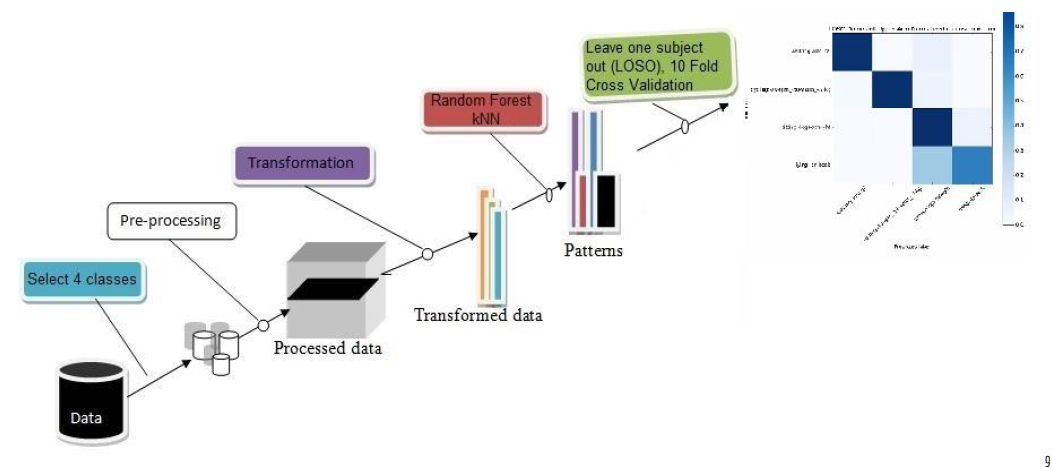
Language: Python

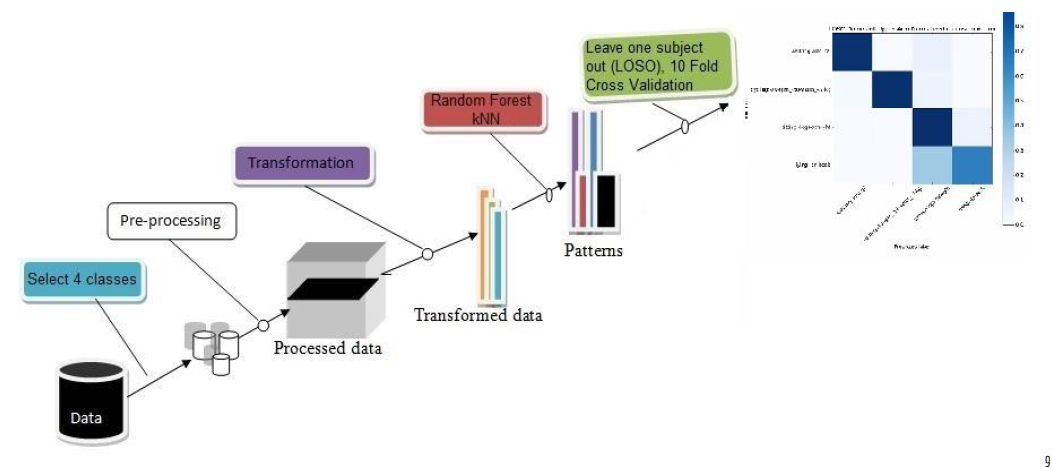
Tools: IPython, PyCharm

Libraries: Scikit-Learn/SciPy, NumPy, Pandas, Matplotlib

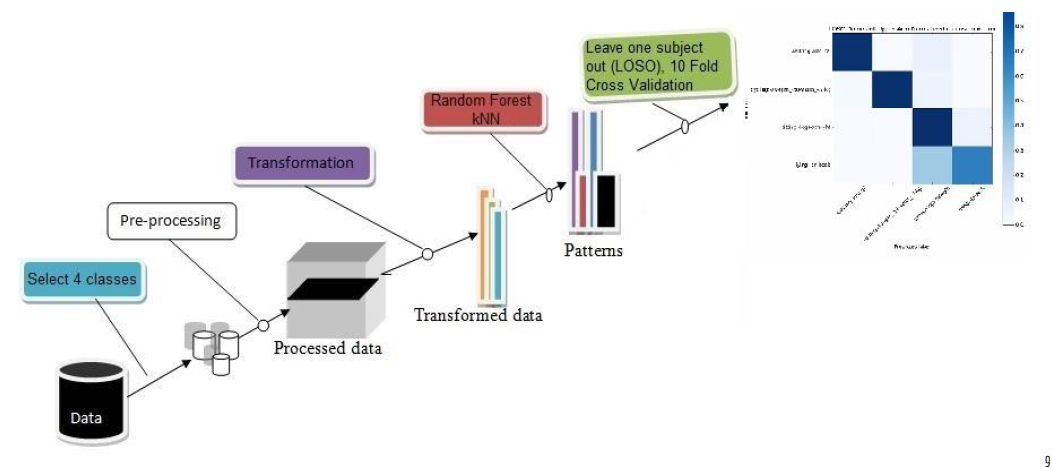
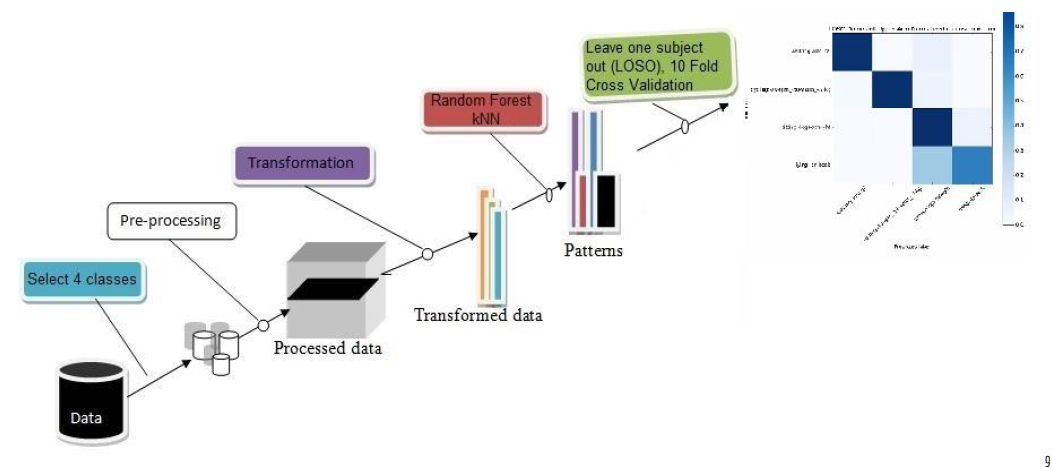
**PROPOSED NOVELITY OR IMPROVEMENT IN TERMS OF ALGORITHM OR NEW FEATURES**

EXISTING ALOGIRTHM:





PROPOSED ARCHITECTURE:

LOSO (LEAVE-ONE -SUBJECT-OUT) VALIDATION

**ANALYSIS OF SET OF EXPERIMENTS PEFORMED**

Dataset: We have raw data set from 33 participants. For each participant

we have following files: -

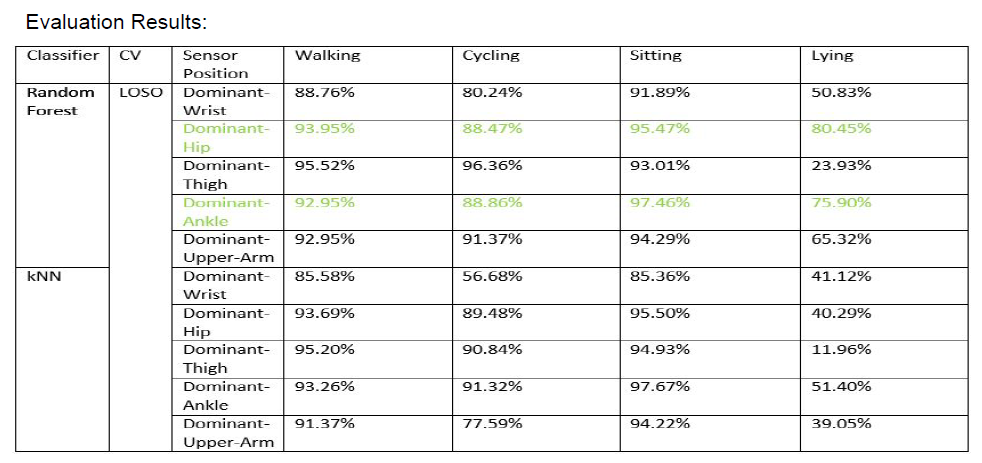
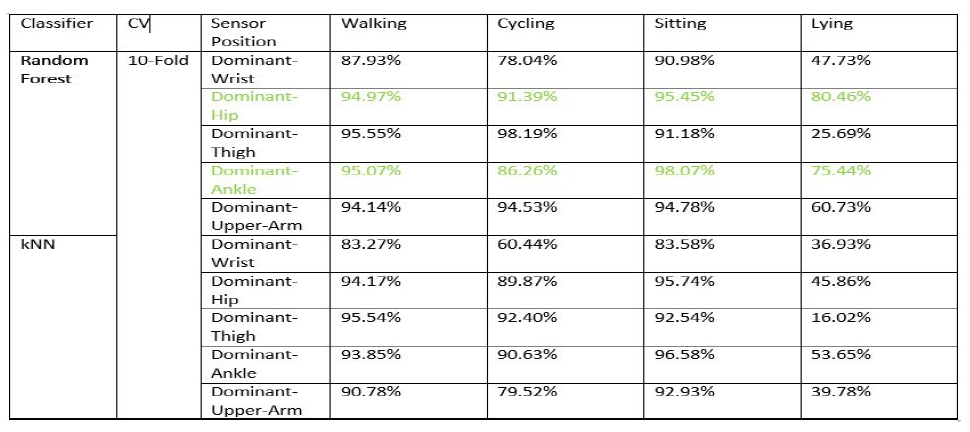
○ Annotations.csv

○ Wocket.csv (A total of 5 files for each sensor location)

Demographics: 33 participants

○ 11 Male, 22 Female, age :18-75, height: 168.5 +/- 9.3cm, weight: 70.0 +/- 15.6 kg

RESULT OF EXPERIMENT FOR LOSO:

RESULT OF EXPERIMEN FOR KNN:

COMBINING DATA FROM THE TOP TWO PLACEMENT SITES, DOMINANT-HIP AND DOMINANT-ANKLE

