

# Human Activity Recognition using Smart-phones

Name:	<b>Jishnu P</b>
Registration No./Roll No.:	19150
Institute/University Name:	IISER Bhopal
Program/Stream:	Physics department
Problem Release date:	January 19, 2022
Date of Submission:	April 24,2022

## 1 Introduction

The objective of the project is to develop supervised machine learning frameworks for the multi-class classification of human activities into one of the six activities performed. They are WALKING, SITTING, STANDING, LAYING, WALKING UPSTAIRS, WALKING DOWNSTAIRS. The Data is collected from 30 subjects using smart phone sensor's accelerometer and gyroscope to record 3 axial linear acceleration and 3 axial angular velocities. By analysing the data's from the sensor's we want to prognosticate what kind of activity is done by the user.

The dataset given for the project is traindata.csv, testdata.csv, and trainlabels.csv. To know the size of the dataset, we use shade method. I found 8239 rows x 563 columns for the traindata.csv and 2060 rows x 563 columns for testdata.csv. The data being multi-dimensional with 561 variables causes so much complications and makes it very difficult to handle and also effect the machine learning training adversely. MRMR is a feature selection algorithm which ensures maximum relevance and minimum redundancy. 1.

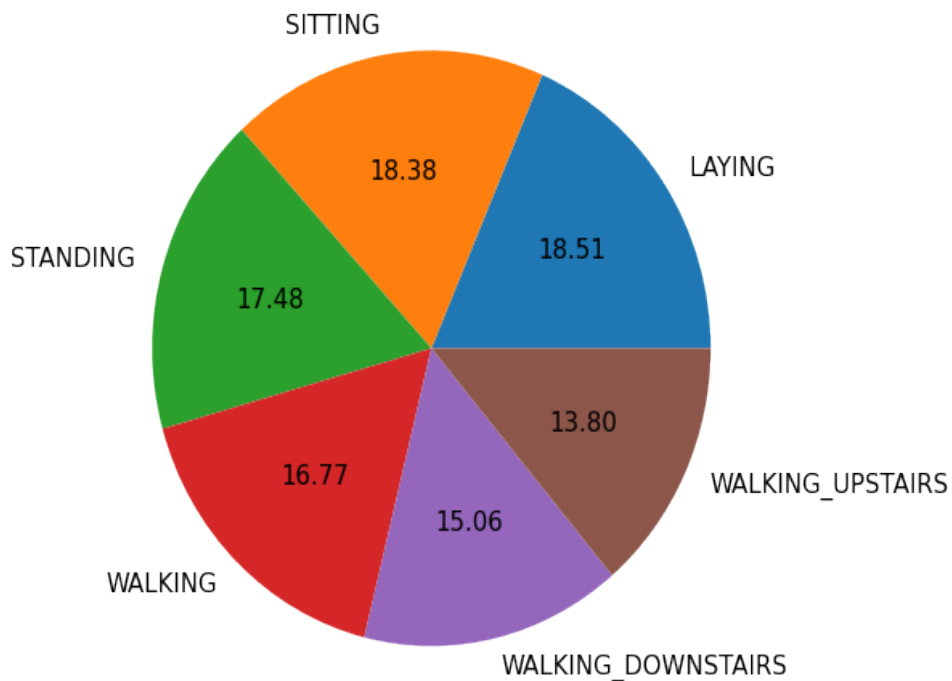


Figure 1: Overview of Corpus

## 2 Methods

The dataset given for the project is traindata.csv, testdata.csv, and trainlabels.csv. To know the size of the dataset, we use shade method. I found 8239 rows x 563 columns for the traindata.csv and 2060 rows x 563 columns for testdata.csv. By running the code, we found that there is non empty set in the data. To check weather the data is biased or not ,we count the number of activities and found that data is uniformly distributed.

The Dataset is slit into training and testing data using traintestspllit. for classification we use many classification techniques with every features Random Forest classifier, logistic regression classifier, KNeighborsClassifier, DecisionTreeClassifier, GaussianNB, support vector machine.The the classifiers are selected for parameter tuning then classifiers shown change in accuracy. For the feature section technique we use MRMR is a feature selection algorithm which ensures maximum relevance and minimum redundancy. it will produce a list of best features. then the classifiers are evaluated with 100,150,200,250,300,350,400,450,500 best features and there confusion matrix and accuracy are noted.

## 3 Evaluation Criteria

Precision quantifies the number of positive class predictions that actually belong to the positive class.

Recall quantifies the number of positive class predictions made out of all positive examples in the dataset.

F-Measure provides a single score that balances both the concerns of precision and recall in one number.

In Micro-average method, you sum up the individual true positives, false positives, and false negatives of the system for different sets and the apply them to get the statistics.

Macro-average Method is straight forward. Just take the average of the precision and recall of the system on different sets

## 4 Analysis of Results

With an accuracy of 0.99, rigorous model evaluation revealed that logistic regression was the highest performing model among all others. Out of 563 features, 400 received the highest score, considerably reducing overfitting in our data.

## 5 Discussions and Conclusion

Project