**Project Report**

**Title: Classification of human activities dataset using Weka tool.**

**Guided By: Submitted By:**

**Dr. Ankush Mittal Mohit Wadhwa**

**Graphic Era Deemed University**

**Contents:**

1. Introduction
2. Dataset
3. Title
4. Contributors
5. Sources
6. Dataset Information
7. Attribute Information
8. Feature Information
9. Dataset Importance
10. Feature Evaluation
11. Classification
12. Without Feature Selection
13. Using J48 Algorithm
14. Using Random Forest Algorithm
15. Using Naïve Bayes Algorithm
16. After Feature Selection using CFS
17. Using J48 Algorithm
18. Using Random Forest Algorithm
19. Using Naïve Bayes Algorithm
20. After Feature Selection using PCA
21. Using J48 Algorithm
22. Using Random Forest Algorithm
23. Using Naïve Bayes Algorithm
24. Comparative Analysis
25. Conclusion
26. **INTRODUCTION:**

****

Human activity recognition through smart phones refers to analyzing the day to day activities like sitting, standing, walking etc with the help of smart phones. Human activity recognition is an important yet challenging research area with many applications in healthcare, smart environments, and homeland security. Computer vision-based techniques have widely been used for human activity tracking, but they mostly require infrastructure support, for example, installation of video cameras in the monitoring areas. Alternatively, a more efficient approach is to process the data from inertial measurement unit sensors worn on a user’s body or built in a user’s smart phone to track his or her motion. We aim to develop a model that is capable of recognizing multiple sets of daily activities under real-world conditions, using data collected by a single tri-axial accelerometer built into a cell phone (in our study, an Android smart phone). A tri-axial accelerometer is a sensor that returns an estimate of acceleration along the x, y and z axes from which velocity and displacement can also be estimated. Activity recognition is formulated as a supervised classification problem, whose training data is obtained via an experiment having human subjects perform each of the activities.

1. **DATASET:**
2. **TITLE:**

### Human Activity Recognition Using Smartphone’s Dataset (Version 1.0)

### CONTRIBUTORS:

### Jorge L. Reyes-Ortiz(1,2), Davide Anguita(1), Alessandro Ghio(1), Luca Oneto(1) and Xavier Parra(2)

### 1 - Smartlab - Non-Linear Complex Systems Laboratory

### DITEN - Università degli Studi di Genova, Genoa (I-16145), Italy.

### 2 - CETpD - Technical Research Centre for Dependency Care and Autonomous Living

### SOURCES:

### The dataset has been collected from UCI Machine Learning Repository. The data can be found at <https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones>

### DATA SET INFORMATION:

### The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities:

### WALKING

### WALKING\_UPSTAIRS

### WALKING\_DOWNSTAIRS

### SITTING

### STANDING

### LAYING

### 

### Image demonstrating various activities observed while wearing a Smart phone

### Above activities were performed wearing a smart phone (Samsung Galaxy S II) on the waist. Using its embedded accelerometer and gyroscope, we captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. The experiments have been video-recorded to label the data manually. The obtained dataset has been randomly partitioned into two sets, where 70% of the volunteers were selected for generating the training data and 30% the test data.

### The sensor signals (accelerometer and gyroscope) were pre-processed by applying noise filters and then sampled in fixed-width sliding windows of 2.56 sec and 50% overlap (128 readings/window). The sensor acceleration signal, which has gravitational and body motion components, was separated using a Butterworth low-pass filter into body acceleration and gravity. The gravitational force is assumed to have only low frequency components; therefore a filter with 0.3 Hz cutoff frequency was used. From each window, a vector of features was obtained by calculating variables from the time and frequency domain.

1. **ATTRIBUTE INFORMATION:**

For each record in the dataset it is provided:

* Tri-axial acceleration from the accelerometer (total acceleration) and the estimated body acceleration.
* Tri-axial Angular velocity from the gyroscope.
* A 561-feature vector with time and frequency domain variables.
* Its activity label.
* An identifier of the subject who carried out the experiment.

1. **FEATURE INFORMATION:**

* Features are normalized and bounded within [-1, 1].
* Each feature vector is a row on the text file.
* The units used for the accelerations (total and body) are 'g's (gravity of earth -> 9.80665 m/seg2).
* The gyroscope units are rad/seg.
* A video of the experiment including an example of the 6 recorded activities with one of the participants can be seen in the following link: http://www.youtube.com/watch?v=XOEN9W05\_4A

1. **DATASET IMPORTANCE:**

Human Activity Recognition System or also commonly called HAR System aims to identify human activities. The identification process can be done by manipulating the information obtained from the surrounding environment or from sensors attached to human body. HAR system that uses sensors mounted on smart phones, such as accelerometer and gyroscope, has the advantage that users do not need to add additional sensors attached on their body. However, the use of smart phones has shortcomings in terms of computing a complex algorithm in real-time. In terms of power supply smart phones only have little capacity and the ability of the processor are more limited than computers in general. Therefore we need to find a way to simplify the computing process by reducing the features from the dataset used using feature selection algorithms.

1. **FEATURE EVALUATION:**

The features selected for this database come from the accelerometer and gyroscope 3-axial raw signals tAcc-XYZ and tGyro-XYZ. These time domain signals (prefix 't' to denote time) were captured at a constant rate of 50 Hz. Then they were filtered using a median filter and a 3rd order low pass Butterworth filter with a corner frequency of 20 Hz to remove noise. Similarly, the acceleration signal was then separated into body and gravity acceleration signals (tBodyAcc-XYZ and tGravityAcc-XYZ) using another low pass Butterworth filter with a corner frequency of 0.3 Hz.

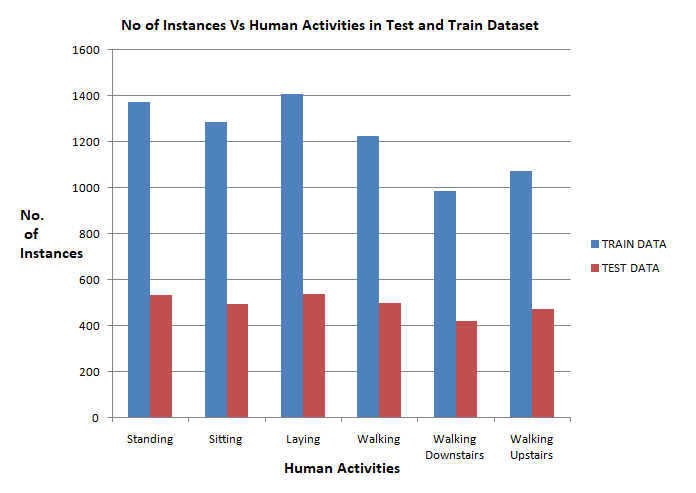
Subsequently, the body linear acceleration and angular velocity were derived in time to obtain Jerk signals (tBodyAccJerk-XYZ and tBodyGyroJerk-XYZ). Also the magnitude of these three-dimensional signals were calculated using the Euclidean norm (tBodyAccMag, tGravityAccMag, tBodyAccJerkMag, tBodyGyroMag, tBodyGyroJerkMag).

Finally a Fast Fourier Transform (FFT) was applied to some of these signals producing fBodyAcc-XYZ, fBodyAccJerk-XYZ, fBodyGyro-XYZ, fBodyAccJerkMag, fBodyGyroMag, fBodyGyroJerkMag. (Note the 'f' to indicate frequency domain signals).

These signals were used to estimate variables of the feature vector for each pattern:

'-XYZ' is used to denote 3-axial signals in the X, Y and Z directions.

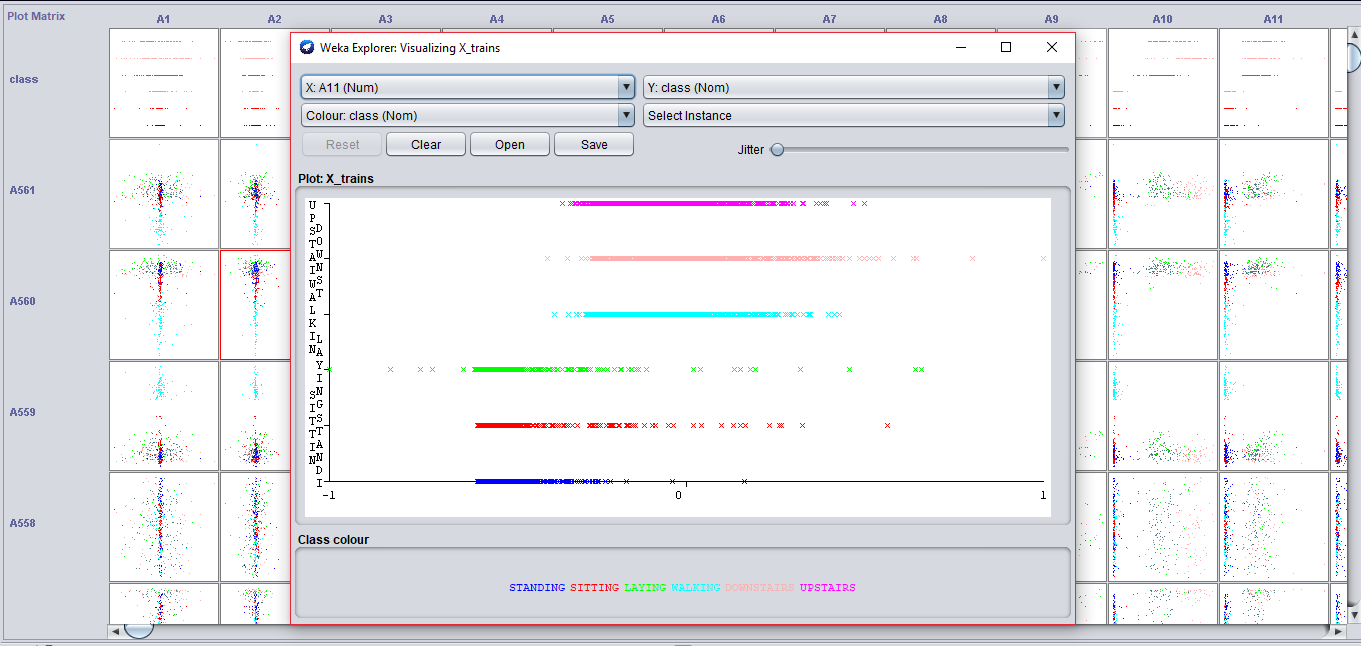
* + - tBodyAcc-XYZ
    - tGravityAcc-XYZ
    - tBodyAccJerk-XYZ
    - tBodyGyro-XYZ
    - tBodyGyroJerk-XYZ
    - tBodyAccMag
    - tGravityAccMag
    - tBodyAccJerkMag
    - tBodyGyroMag
    - tBodyGyroJerkMag
    - fBodyAcc-XYZ
    - fBodyAccJerk-XYZ
    - fBodyGyro-XYZ
    - fBodyAccMag
    - fBodyAccJerkMag
    - fBodyGyroMag
    - fBodyGyroJerkMag



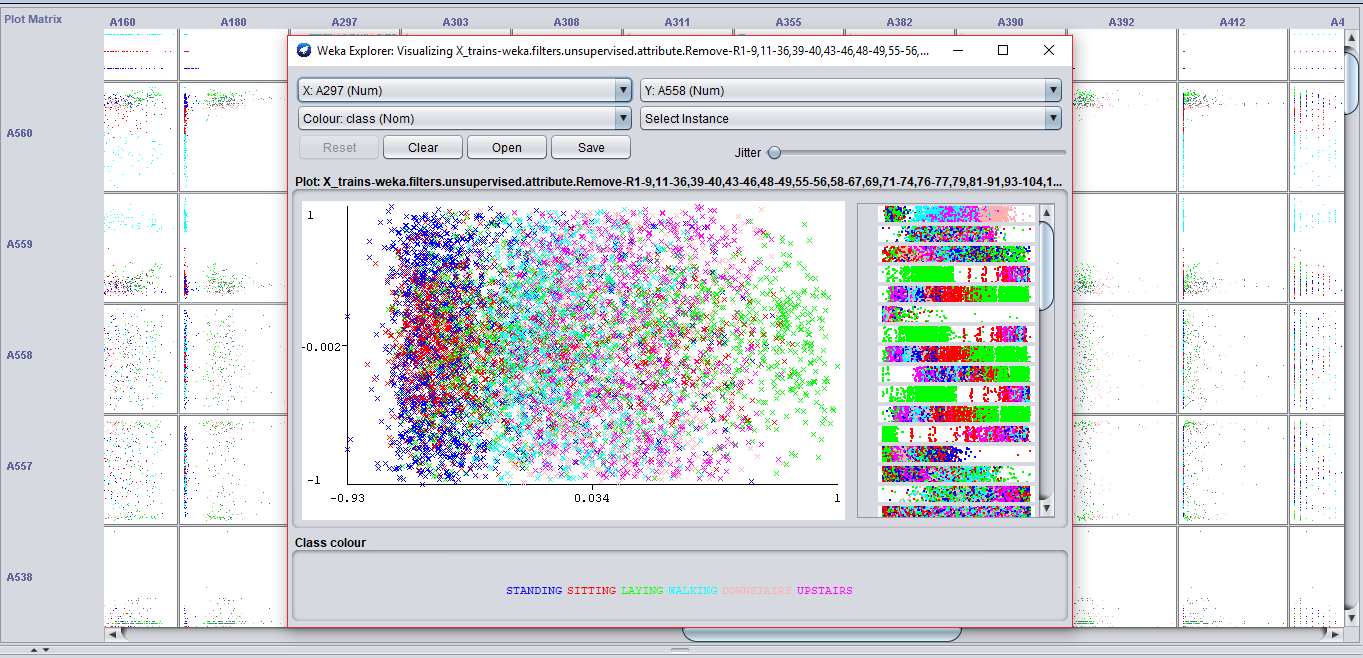
1. **Classification:**

Classification refers to providing a predefined class to the data in the data set . Classification is done in association with attributes (features).Thus, feature selection is an important step in classification. Various results are shown below demonstrating the use of various selection algorithms on different classification algorithms. We can determine if a feature is useful for us or not by visualizing the graph of that feature . If in the graph we are able to visualize different classes then we can say that the feature plays an essential role in determining the accuracy of classification . But , if the graph is scattered in such a way that it is difficult to visualize various classes then we may say that the corresponding attribute is not a useful attribute , thus could be eliminated.

The below graph shows visualization of a useful attribute:



The below graph shows visualization of a non-useful attribute:



1. **Without Feature selection:**

|  |  |
| --- | --- |
| Algorithm Used | J48 pruned tree |
| Instances | 7352 |
| Attributes | 562 |

|  |  |  |
| --- | --- | --- |
| Correctly Classified Instances | 2461 |  |
| Incorrectly Classified Instances | 486 |  |
| Kappa statistic | 0.8018 | |
| Mean absolute error | 0.055 | |
| Root mean squared error | 0.2298 | |
| Relative absolute error | 19.8548 % | |
| Root relative squared error | 61.6738 % | |
| Total Number of Instances | 2947 | |

**Accuracy: 83.5087 %**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
| 0.874 | 0.039 | 0.832 | 0.874 | 0.852 | 0.819 | 0.957 | 0.799 | STANDING |
| 0.809 | 0.027 | 0.856 | 0.809 | 0.831 | 0.799 | 0.907 | 0.758 | SITTING |
| 1.000 | 0.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | LAYING |
| 0.768 | 0.032 | 0.830 | 0.768 | 0.798 | 0.760 | 0.954 | 0.828 | WALKING |
| 0.771 | 0.026 | 0.831 | 0.771 | 0.800 | 0.769 | 0.865 | 0.687 | DOWNSTAIRS |
| 0.758 | 0.073 | 0.664 | 0.758 | 0.708 | 0.650 | 0.883 | 0.633 | UPSTAIRS |
| W Ag. 0.835 | 0.032 | 0.839 | 0.835 | 0.836 | 0.805 | 0.931 | 0.791 | |

**Confusion Matrix:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **a** | **B** | **c** | **D** | **e** | **f** | **classified as** |
| 465 | 67 | 0 | 0 | 0 | 0 | **a = STANDING** |
| 94 | 397 | 0 | 0 | 0 | 0 | **b = SITTING** |
| 0 | 0 | 537 | 0 | 0 | 0 | **c = LAYING** |
| 0 | 0 | 0 | 381 | 13 | 102 | **d = WALKING** |
| 0 | 0 | 0 | 17 | 324 | 79 | **e = DOWNSTAIRS** |
| 0 | 0 | 0 | 61 | 53 | 357 | **f = UPSTAIRS** |

**Analysis Drawn:** On using J48 pruned tree, without any feature selection, we see that out of 2947, 2461 instances got correctly classified and 486 instances got wrongly classified leaving an accuracy of  **83.5087%** . This is the best accuracy we have got so far. We shall try to improve the accuracy in further experimentations.

|  |  |
| --- | --- |
| Algorithm Used | Random Forest |
| Instances | 7352 |
| Attributes | 562 |

|  |  |  |
| --- | --- | --- |
| Correctly Classified Instances | 2764 |  |
| Incorrectly Classified Instances | 183 |  |
| Kappa statistic | 0.9253 | |
| Mean absolute error | 0.0738 | |
| Root mean squared error | 0.1518 | |
| Relative absolute error | 26.6341 % | |
| Root relative squared error | 40.7516 % | |
| Total Number of Instances | 2947 | |

**Accuracy: 93.7903 %**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
| 0.970 | 0.021 | 0.910 | 0.970 | 0.939 | 0.926 | 0.997 | 0.988 | STANDING |
| 0.896 | 0.007 | 0.965 | 0.896 | 0.929 | 0.917 | 0.997 | 0.987 | SITTING |
| 1.000 | 0.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | LAYING |
| 0.974 | 0.023 | 0.894 | 0.974 | 0.932 | 0.919 | 0.998 | 0.989 | WALKING |
| 0.838 | 0.004 | 0.970 | 0.838 | 0.899 | 0.887 | 0.994 | 0.971 | DOWNSTAIRS |
| 0.926 | 0.019 | 0.901 | 0.926 | 0.913 | 0.896 | 0.993 | 0.957 | UPSTAIRS |
| W Ag. 0.938 | 0.013 | 0.940 | 0.938 | 0.938 | 0.926 | 0.997 | 0.983 | |

**Confusion Matrix:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **a** | **B** | **c** | **D** | **e** | **f** | **classified as** |
| 516 | 16 | 0 | 0 | 0 | 0 | **a = STANDING** |
| 51 | 440 | 0 | 0 | 0 | 0 | **b = SITTING** |
| 0 | 0 | 537 | 0 | 0 | 0 | **c = LAYING** |
| 0 | 0 | 0 | 483 | 5 | 8 | **d = WALKING** |
| 0 | 0 | 0 | 28 | 352 | 40 | **e = DOWNSTAIRS** |
| 0 | 0 | 0 | 29 | 6 | 436 | **f = UPSTAIRS** |

**Analysis Drawn:** On using Random Forest, without any feature selection, we see that out of 2947, 2764 instances got correctly classified and 183 instances got wrongly classified leaving an accuracy of  **93.7903 %**. This is the best accuracy we have got so far. We can clearly see that Random Forest algorithm has surpassed J48 in terms of accuracy for this data set. We shall try to improve the accuracy in further experimentations.

|  |  |
| --- | --- |
| Algorithm Used | Naïve Bayes |
| Instances | 7352 |
| Attributes | 562 |

|  |  |  |
| --- | --- | --- |
| Correctly Classified Instances | 2311 |  |
| Incorrectly Classified Instances | 636 |  |
| Kappa statistic | 0.7404 | |
| Mean absolute error | 0.0717 | |
| Root mean squared error | 0.2657 | |
| Relative absolute error | 25.855 % | |
| Root relative squared error | 71.3233 % | |
| Total Number of Instances | 2947 | |

**Accuracy: 78.4187%**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
| 0.912 | 0.082 | 0.709 | 0.912 | 0.798 | 0.755 | 0.945 | 0.800 | STANDING |
| 0.558 | 0.043 | 0.721 | 0.558 | 0.629 | 0.572 | 0.926 | 0.743 | SITTING |
| 0.825 | 0.014 | 0.929 | 0.825 | 0.874 | 0.850 | 0.989 | 0.922 | LAYING |
| 0.815 | 0.037 | 0.816 | 0.815 | 0.815 | 0.778 | 0.962 | 0.810 | WALKING |
| 0.612 | 0.025 | 0.806 | 0.612 | 0.696 | 0.661 | 0.952 | 0.755 | DOWNSTAIRS |
| 0.951 | 0.058 | 0.757 | 0.951 | 0.843 | 0.817 | 0.973 | 0.792 | UPSTAIRS |
| W Ag.0.784 | 0.044 | 0.790 | 0.784 | 0.779 | 0.742 | 0.958 | 0.807 |  |

**Confusion Matrix:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **a** | **B** | **c** | **D** | **e** | **f** | **classified as** |
| 485 | 14 | 20 | 0 | 0 | 13 | **a = STANDING** |
| 199 | 274 | 14 | 0 | 0 | 4 | **b = SITTING** |
| 0 | 92 | 443 | 0 | 0 | 2 | **c = LAYING** |
| 0 | 0 | 0 | 404 | 49 | 43 | **d = WALKING** |
| 0 | 0 | 0 | 81 | 257 | 82 | **e = DOWNSTAIRS** |
| 0 | 0 | 0 | 10 | 13 | 448 | **f = UPSTAIRS** |

**Analysis Drawn:** On using Naïve Bayes, without any feature selection, we see that out of 2947, 2311 instances got correctly classified and 636 instances got wrongly classified leaving an accuracy of  **78.4187 %**. We can clearly see that Random Forest algorithm has surpassed J48 and Naïve Bayes in terms of accuracy for this data set. We shall try to improve the accuracy in further experimentations.

1. **By selecting the features with the help of Co-relation based Feature Selection (CFS) algorithm.**

|  |  |
| --- | --- |
| Total Instances | 7352 |
| Total Attributes before using PCA | 562 |

|  |  |
| --- | --- |
| Total Instances | 7352 |
| Total Attributes after using PCA | 47 |

**Selected Attributes:**

10,37,38,41,42,47,50,51,52,53,54,57,68,70,75,78,80,92,105,108,121,159,160,180,297,303,308,311,355,382,390,392,412,449,450,451,452,490,503,504,518,519,538,557,558,559,560 : Total : 47

|  |  |
| --- | --- |
| Algorithm Used | J48 |
| Instances | 7352 |
| Attributes | 562 |

|  |  |  |
| --- | --- | --- |
| Correctly Classified Instances | 2498 |  |
| Incorrectly Classified Instances | 449 |  |
| Kappa statistic | 0.8167 | |
| Mean absolute error | 0.0516 | |
| Root mean squared error | 0.2228 | |
| Relative absolute error | 18.6265 % | |
| Root relative squared error | 59.8141 % | |
| Total Number of Instances | 2947 | |

**Accuracy:** 84.7642 %

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TP  Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC  Area | PRC  Area | Class |
| 0.852 | 0.049 | 0.793 | 0.852 | 0.821 | 0.781 | 0.935 | 0.731 | STANDING |
| 0.760 | 0.032 | 0.825 | 0.760 | 0.791 | 0.752 | 0.862 | 0.672 | SITTING |
| 1.000 | 0.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | LAYING |
| 0.915 | 0.046 | 0.802 | 0.915 | 0.855 | 0.826 | 0.938 | 0.798 | WALKING |
| 0.729 | 0.012 | 0.908 | 0.729 | 0.808 | 0.787 | 0.890 | 0.735 | DOWNSTAIRS |
| 0.796 | 0.044 | 0.775 | 0.796 | 0.785 | 0.744 | 0.905 | 0.654 | UPSTAIRS |
| W Ag.0.848 | 0.031 | 0.851 | 0.848 | 0.847 | 0.819 | 0.924 | 0.770 |  |

**Confusion Matrix:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **a** | **b** | **c** | **D** | **e** | **f** | **classified as** |
| 453 | 79 | 0 | 0 | 0 | 0 | **a = STANDING** |
| 118 | 373 | 0 | 0 | 0 | 0 | **b = SITTING** |
| 0 | 0 | 537 | 0 | 0 | 0 | **c = LAYING** |
| 0 | 0 | 0 | 454 | 19 | 23 | **d = WALKING** |
| 0 | 0 | 0 | 28 | 306 | 86 | **e = DOWNSTAIRS** |
| 0 | 0 | 0 | 84 | 12 | 375 | **f = UPSTAIRS** |

**Analysis Drawn:** On using J48 pruned tree, with feature selection through Co-relation based Feature Selection (CFS) algorithm selecting a total of 47 features out of 562, we see that out of 2947, 2498 instances got correctly classified and 449 instances got wrongly classified leaving an accuracy of  **84.7642 %**.We shall try to improve the accuracy in further experimentations.

|  |  |
| --- | --- |
| Algorithm Used | Random Forest |
| Instances | 7352 |
| Attributes | 48 |

|  |  |  |
| --- | --- | --- |
| Correctly Classified Instances | 2631 |  |
| Incorrectly Classified Instances | 316 |  |
| Kappa statistic | 0.8711 | |
| Mean absolute error | 0.0583 | |
| Root mean squared error | 0.1624 | |
| Relative absolute error | 21.0285 % | |
| Root relative squared error | 43.5921% | |
| Total Number of Instances | 2947 | |

**Accuracy: 89.2772%**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
| 0.868 | 0.044 | 0.812 | 0.868 | 0.839 | 0.803 | 0.984 | 0.919 | STANDING |
| 0.782 | 0.029 | 0.846 | 0.782 | 0.813 | 0.778 | 0.982 | 0.918 | SITTING |
| 1.000 | 0.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | LAYING |
| 0.972 | 0.022 | 0.898 | 0.972 | 0.933 | 0.920 | 0.997 | 0.989 | WALKING |
| 0.840 | 0.006 | 0.957 | 0.840 | 0.895 | 0.881 | 0.993 | 0.967 | DOWNSTAIRS |
| 0.877 | 0.027 | 0.859 | 0.877 | 0.868 | 0.842 | 0.989 | 0.938 | UPSTAIRS |
| WAvg.0.893 | 0.022 | 0.894 | 0.893 | 0.892 | 0.872 | 0.991 | 0.955 |  |

**Confusion Matrix:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **a** | **b** | **c** | **D** | **E** | **f** | **classified as** |
| 462 | 70 | 0 | 0 | 0 | 0 | **a = STANDING** |
| 107 | 384 | 0 | 0 | 0 | 0 | **b = SITTING** |
| 0 | 0 | 537 | 0 | 0 | 0 | **c = LAYING** |
| 0 | 0 | 0 | 482 | 2 | 12 | **d = WALKING** |
| 0 | 0 | 0 | 11 | 353 | 56 | **e = DOWNSTAIRS** |
| 0 | 0 | 0 | 44 | 14 | 413 | **f = UPSTAIRS** |

**Analysis Drawn:** On using Random Forest, with feature selection through Co-relation based Feature Selection (CFS) algorithm selecting a total of 48 features out of 562, we see that out of 2947, 2631 instances got correctly classified and 316 instances got wrongly classified leaving an accuracy of  **89.2772%** . We can clearly see that Random Forest algorithm has surpassed J48 in terms of accuracy for this data set. We shall try to improve the accuracy in further experimentations.

|  |  |
| --- | --- |
| Algorithm Used | Naïve Bayes |
| Instances | 7352 |
| Attributes | 48 |

|  |  |  |
| --- | --- | --- |
| Correctly Classified Instances | 2546 |  |
| Incorrectly Classified Instances | 401 |  |
| Kappa statistic | 0.8363 | |
| Mean absolute error | 0.0458 | |
| Root mean squared error | 0.2063 | |
| Relative absolute error | 16.5321 % | |
| Root relative squared error | 55.3689% | |
| Total Number of Instances | 2947 | |

**Accuracy: 86.3929%**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | ROC Area | Class |
| 0.932 | 0.072 | 0.739 | 0.932 | 0.825 | 0.789 | 0.980 | 0.915 | STANDING |
| 0.633 | 0.010 | 0.928 | 0.633 | 0.753 | 0.732 | 0.980 | 0.912 | SITTING |
| 1.000 | 0.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | LAYING |
| 0.948 | 0.040 | 0.827 | 0.948 | 0.883 | 0.861 | 0.992 | 0.967 | WALKING |
| 0.769 | 0.013 | 0.905 | 0.769 | 0.831 | 0.810 | 0.985 | 0.924 | DOWNSTAIRS |
| 0.868 | 0.028 | 0.854 | 0.868 | 0.861 | 0.834 | 0.987 | 0.912 | UPSTAIRS |
| W Avg.0.864 | 0.028 | 0.875 | 0.864 | 0.861 | 0.840 | 0.987 | 0.940 |  |

**Confusion Matrix:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **a** | **b** | **c** | **D** | **E** | **f** | **classified as** |
| 496 | 24 | 0 | 5 | 0 | 7 | **a = STANDING** |
| 175 | 311 | 0 | 0 | 0 | 5 | **b = SITTING** |
| 0 | 0 | 537 | 0 | 0 | 0 | **c = LAYING** |
| 0 | 0 | 0 | 470 | 23 | 3 | **d = WALKING** |
| 0 | 0 | 0 | 42 | 323 | 55 | **e = DOWNSTAIRS** |
| 0 | 0 | 0 | 51 | 11 | 409 | **f = UPSTAIRS** |

**Analysis Drawn:** On using Naïve Bayes, with feature selection through Co-relation based Feature Selection (CFS) algorithm selecting a total of 48 features out of 562, we see that out of 2947, 2546 instances got correctly classified and 401 instances got wrongly classified leaving an accuracy of  **86.3929%.**We can clearly see that Random Forest algorithm has surpassed J48 and Naïve Bayes in terms of accuracy for this data set. We shall try to improve the accuracy in further experimentations.

**3-Principal Component:**

|  |  |
| --- | --- |
| Total Instances | 7352 |
| Total Attributes before using PCA | 562 |

|  |  |
| --- | --- |
| Total Instances | 7352 |
| Total Attributes after using PCA | 102 |

**Selected Attributes:**

1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97,98,99,100,101,102 :

Total : 102

|  |  |
| --- | --- |
| Algorithm Used | J48 |
| Instances | 7352 |
| Attributes | 102 |

|  |  |  |
| --- | --- | --- |
| Correctly Classified Instances | 2389 |  |
| Incorrectly Classified Instances | 558 |  |
| Kappa statistic | 0.7723 | |
| Mean absolute error | 0.0632 | |
| Root mean squared error | 0.2477 | |
| Relative absolute error | 22.8053 % | |
| Root relative squared error | 66.4907% | |
| Total Number of Instances | 2947 | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
| 0.789 | 0.057 | 0.753 | 0.789 | 0.771 | 0.719 | 0.904 | 0.703 | STANDING |
| 0.719 | 0.046 | 0.759 | 0.719 | 0.738 | 0.688 | 0.862 | 0.605 | SITTING |
| 1.000 | 0.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | LAYING |
| 0.881 | 0.062 | 0.741 | 0.881 | 0.805 | 0.765 | 0.892 | 0.619 | WALKING |
| 0.779 | 0.023 | 0.847 | 0.779 | 0.811 | 0.783 | 0.878 | 0.750 | DOWNSTAIRS |
| 0.669 | 0.039 | 0.766 | 0.669 | 0.714 | 0.666 | 0.838 | 0.584 | UPSTAIRS |
| W Avg.0.811 | 0.038 | 0.812 | 0.811 | 0.810 | 0.774 | 0.898 | 0.714 |  |

**Accuracy: 81.0655%**

**Confusion Matrix:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **a** | **b** | **c** | **D** | **E** | **f** | **classified as** |
| 420 | 112 | 0 | 0 | 0 | 0 | **a = STANDING** |
| 138 | 353 | 0 | 0 | 0 | 0 | **b = SITTING** |
| 0 | 0 | 537 | 0 | 0 | 0 | **c = LAYING** |
| 0 | 0 | 0 | 437 | 24 | 35 | **d = WALKING** |
| 0 | 0 | 0 | 32 | 327 | 61 | **e = DOWNSTAIRS** |
| 0 | 0 | 0 | 121 | 35 | 315 | **f = UPSTAIRS** |

**Analysis Drawn:** On using J48 pruned tree, with feature selection through Principal Component algorithm selecting a total of 102 features out of 562, we see that out of 2947, 2389 instances got correctly classified and 558 instances got wrongly classified leaving an accuracy of  **81.0655%.**We shall try to improve the accuracy in further experimentations.

|  |  |
| --- | --- |
| Algorithm Used | Random Forest |
| Instances | 7352 |
| Attributes | 102 |

|  |  |  |
| --- | --- | --- |
| Correctly Classified Instances | 2601 |  |
| Incorrectly Classified Instances | 346 |  |
| Kappa statistic | 0.8588 | |
| Mean absolute error | 0.0704 | |
| Root mean squared error | 0.1752 | |
| Relative absolute error | 25.3792 % | |
| Root relative squared error | 47.0306% | |
| Total Number of Instances | 2947 | |

**Accuracy: 88.2592%**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
| 0.829 | 0.035 | 0.838 | 0.829 | 0.834 | 0.797 | 0.980 | 0.899 | STANDING |
| 0.827 | 0.037 | 0.817 | 0.827 | 0.822 | 0.786 | 0.978 | 0.897 | SITTING |
| 1.000 | 0.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | LAYING |
| 0.952 | 0.036 | 0.841 | 0.952 | 0.893 | 0.872 | 0.994 | 0.972 | WALKING |
| 0.817 | 0.004 | 0.974 | 0.817 | 0.889 | 0.876 | 0.989 | 0.957 | DOWNSTAIRS |
| 0.854 | 0.029 | 0.848 | 0.854 | 0.851 | 0.822 | 0.985 | 0.918 | UPSTAIRS |
| W Avg.0.883 | 0.024 | 0.886 | 0.883 | 0.883 | 0.860 | 0.988 | 0.940 |  |

**Confusion Matrix**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **a** | **b** | **c** | **D** | **E** | **f** | **classified as** |
| 441 | 91 | 0 | 0 | 0 | 0 | **a = STANDING** |
| 85 | 406 | 0 | 0 | 0 | 0 | **b = SITTING** |
| 0 | 0 | 537 | 0 | 0 | 0 | **c = LAYING** |
| 0 | 0 | 0 | 472 | 2 | 22 | **d = WALKING** |
| 0 | 0 | 0 | 27 | 343 | 50 | **e = DOWNSTAIRS** |
| 0 | 0 | 0 | 62 | 7 | 402 | **f = UPSTAIRS** |

**Analysis Drawn:** On using Random Forest, with feature selection through Principal Component algorithm selecting a total of 102 features out of 562, we see that out of 2947, 2601 instances got correctly classified and 346 instances got wrongly classified leaving an accuracy of  **88.2592%**. We can clearly see that Random Forest algorithm has surpassed J48 in terms of accuracy for this data set. We shall try to improve the accuracy in further experimentations.

|  |  |
| --- | --- |
| Algorithm Used | NaiveBayes |
| Instances | 7352 |
| Attributes | 102 |

|  |  |  |
| --- | --- | --- |
| Correctly Classified Instances | 2386 |  |
| Incorrectly Classified Instances | 561 |  |
| Kappa statistic | 0.7708 | |
| Mean absolute error | 0.0631 | |
| Root mean squared error | 0.2478 | |
| Relative absolute error | 22.7765 % | |
| Root relative squared error | 66.5072% | |
| Total Number of Instances | 2947 | |

**Accuracy: 80.9637%**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
| 0.949 | 0.101 | 0.674 | 0.949 | 0.788 | 0.749 | 0.966 | 0.842 | STANDING |
| 0.485 | 0.007 | 0.937 | 0.485 | 0.639 | 0.635 | 0.964 | 0.851 | SITTING |
| 0.996 | 0.004 | 0.982 | 0.996 | 0.989 | 0.986 | 0.996 | 0.982 | LAYING |
| 0.881 | 0.065 | 0.733 | 0.881 | 0.800 | 0.760 | 0.975 | 0.913 | WALKING |
| 0.679 | 0.025 | 0.819 | 0.679 | 0.742 | 0.708 | 0.971 | 0.828 | DOWNSTAIRS |
| 0.820 | 0.028 | 0.848 | 0.820 | 0.834 | 0.803 | 0.984 | 0.884 | UPSTAIRS |
| W Avg.0.810 | 0.039 | 0.832 | 0.810 | 0.803 | 0.778 | 0.976 | 0.886 |  |

**Confusion Matrix**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **a** | **b** | **c** | **D** | **E** | **f** | **classified as** |
| 505 | 16 | 5 | 0 | 1 | 5 | **a = STANDING** |
| 244 | 238 | 5 | 0 | 1 | 3 | **b = SITTING** |
| 0 | 0 | 535 | 0 | 0 | 2 | **c = LAYING** |
| 0 | 0 | 0 | 437 | 44 | 15 | **d = WALKING** |
| 0 | 0 | 0 | 91 | 285 | 44 | **e = DOWNSTAIRS** |
| 0 | 0 | 0 | 68 | 17 | 386 | **f = UPSTAIRS** |

**Analysis Drawn:** On using Naïve Bayes with feature selection through Principal Component algorithm selecting a total of 102 features out of 562, we see that out of 2947, 2386 instances got correctly classified and 561 instances got wrongly classified leaving an accuracy of **80.9637%.**We can clearly see that Random Forest algorithm has surpassed J48 and Naïve Bayes in terms of accuracy for this data set. We shall try to improve the accuracy in further experimentations.

1. **CONCLUSION:**

Human activity recognition is an important yet challenging research area with many applications in healthcare, smart environments, and homeland security. We observe that accuracy is very essential for activity recognition.

Through this project we could draw an analysis that without any feature selection algorithm the best results for this data set are provided in the order Random Forest algorithm being the best then J48 pruned tree algorithm and then Naïve Bayes algorithm with lowest accuracy. Using feature selection algorithm we see that through Co-relation based Feature Selection (CFS) algorithm (selecting a total of 48 attributes out of 562) we get better results as compared to the Principal Component algorithm (selecting a total of 102 attributes out of 562), the classification algorithm in increasing order of accuracy being Naïve Bayes algorithm, J48 pruned tree algorithm and Random Forest algorithm.

Thus we can conclude through our observations that the maximum accuracy is obtained in case of Random Forest (93.7903 %). Hence we can say that Random Forest is the best classification algorithm for human activity recognition through smart phones.