Human Activity Model Analyser

[1] Sherry Jain, [2] Chetan Bhargava, [3] Tanvi Jain, [4] Vishakha Tyagi, Mr. Devender Banga

1,2,3,4 Student, Department of Information Technology, Dr. Akhilesh Das Gupta Institute of Technology and Management, Delhi, India.
Assistant Professor, Department of Information Technology, Dr. Akhilesh Das Gupta Institute of Technology and Management, Delhi, India.

ABSTRACT: Human activity recognition has wide applications in medical research and human survey system. This project depicts recognition of Human activities using data generated from user's Smartphone. We have used data available at University of California Machine Learning repository to recognize six human activities. These activities are Standing, Sitting, Laying, Walking, upstairs and walking downstairs. Data is collected from an embedded accelerometer, gyroscope and other sensors of Samsung Galaxy S II Smartphone.

I. INTRODUCTION

Nowadays large amount of data is available everywhere. Therefore, it is very important to analyze this data in order to extract some useful information and to develop an algorithm based on this analysis. This can be achieved through data mining and machine learning. Machine learning is an integral part of artificial intelligence, which is used to design algorithms based on the data trends and historical relationships between data. Machine learning is used in various fields such as bioinformatics, intrusion detection, Information retrieval, game playing, marketing, malware detection, image deconvolution and so on.

II. RELATED WORK

Human activity recognition on smartphone is an active research area. Most of related works focus on analyzing the performance of classification algorithms such as: Random forest,K.Nearest Neighbor Support Vector Machines, logistic regression algorithms, Human Activity Recognition (HAR) aims to identify the actions carried out by a person given a set of observations of him/her and the surrounding environment. Training Model can be accomplished by exploiting the information retrieved from various sources such as environmental or body-worn sensor data. The idea is that once the subject's activity is recognized and

Data is randomly divided into 7:3 ratios to form training and testing dataset respectively. Dimensionality reduction is done using Principal Component Analysis technique. Activity classification is done using Machine Learning models namely Logistic Regression, Random Forest, Support Vector Machine, and K-Nearest Neighbor. We have compared accuracy and performance of these models using accuracy score.

known, an intelligent computer system can then offer assistance. Human activity recognition has wide applications in medical research and human survey system .This project depicts recognition of Human activities using data generated from user's Smartphone. We have used data available at University of California Machine Learning repository to recognize six human activities. These activities are Standing, Sitting, Laying, Walking, upstairs and walking downstairs. Data is collected from an embedded accelerometer, gyroscope and other sensors of Samsung Galaxy S II Smartphone. Data is randomly divided into 7:3 ratios to form training and testing dataset respectively. Dimensionality reduction is done using Principal Component Analysis technique. Activity classification is done using Machine Learning models namely Logistic Regression, Random Forest, Support Vector Machine and K-Nearest Neighbor. We have compared accuracy and performance of these models using accuracy score.

III. CORE TECHNIQUES

1.The KNN (k-nearest neighbors) : The KNN(k-nearest neighbors) algorithm is supervised machine learning algorithm which can be used for both classification and regression problems . In the industry it is widely used for classification algorithms.

How does KNN algorithm work?

(1) Pick a value of k .

(2) Search for the k observation in the training data that are "nearest" to the measurements of the unknown data .

(3) Use the most popular response from the k - nearest neighbor as the predicted response for the unknown data .

(4) Then by using accuracy score we will made the required Predictions .
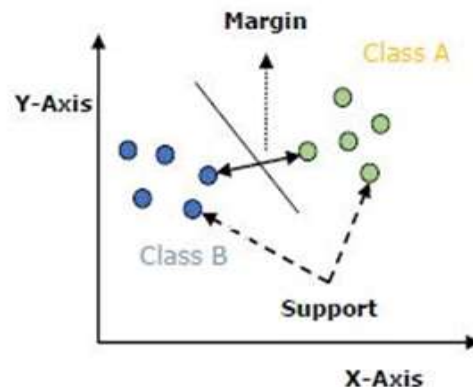
KNN ALGORITHM
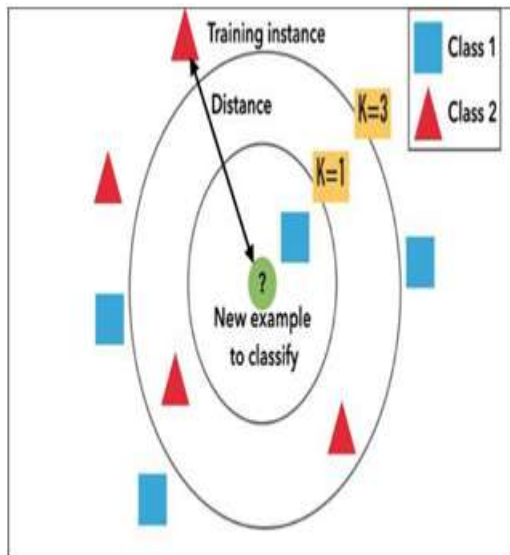
2.Random forest or random decision forest Random forest or random decision forest are an ensemble learning method for both the classification as well as regression. In ransom forest we build multiple decision trees and joins them together to get an accurate prediction. It generally uses the same hyper parameters used by decision tree or a bagging classifier. Versatility is one of the biggest advantage of random forest and we can use this in a lot of different fields like banking , stock market , medicine and e - commerce . It is a great algorithm to train model early in the development process, to see how it behaves and performs and it also very difficult to beat on the basis of performance. Overall , it is a fast flexible and simple tool .

RANDOM FOREST ALGORITHM
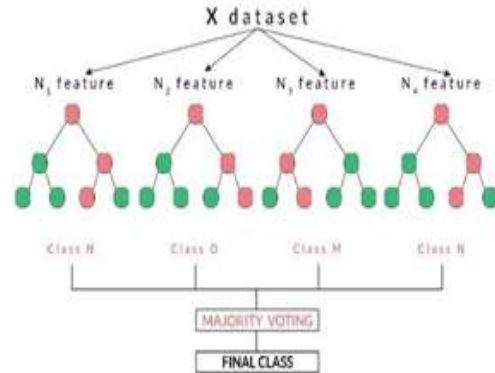
3.SVM : It is differ from all other classification algorithms in the way they select the decision boundary that maximizes the distance from the nearest data points of all the class. The decision boundary create by SVMs is known as the maximum margin classifier or the maximum margin hyper plane. A simple linear SVM classifier work by creating a straight line between two classes. Meaning all the data points on one side of the line will represent category and the data points on the other side of the line will be put in the different category. That mean that there can be an infinite number of lines to choose from. We use SVMs as they helps to find complex relationships between the data without us needing to do a lot of transformations on our own. It is a great option when we are working with small dataset that have thousands of features. They finds more correct result when compared to other algorithm because of the ability to handle small, complex datasets.

SVM ALGORITHM

4.Logistic Regression : It is a supervised learning algorithm used in predicting the probability of a target variable. It is a statistical model which makes use of a logistic function in modelling of a binary

dependent variable. It is estimating and calculating the parameters of a logistic model for regression analysis. It is used test the association of categorical or continuous independent variables having one dichotomous dependent variable. Unlike linear regression analysis in which dependent variable is a continuous variable. The linear regression model works well for regression but fails for classification.Logistic regression is a supervised learning algorithm used in predicting the probability of a target variable. It is a statistical model which makes use of a logistic function in modelling of a binary dependent variable. A logistic function is as follows:

$$\text{logistic}(\eta) = \frac{1}{1 + exp(-\eta)}$$

It is estimating and calculating the parameters of a logistic model for regression analysis. It is used test the association of categorical or continuous independent variables having one dichotomous dependent variable. Unlike linear regression analysis in which dependent variable is a continuous variable. The linear regression model works well for regression but fails for classification.

Logistic Regression Example
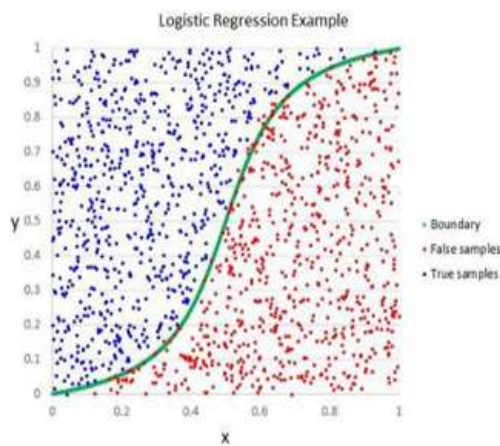
IV. HUMAN ACTIVITY MODEL ANALYSER USING SMARTPHONE DATASET

Our aim is to build a classification model that can identify human fitness activities with a high degree of accuracy. So to do so, first we will do the required cleaning for our dataset and then we will divide our dataset into two parts (training and testing) and by applying different machine learning models with best tuning parameters we will do predictions to see which classification model can identify human activity with a high accuracy rate.

1.Data Collection

We have collected the dataset from https://www.kaggle.com. The Human Activity Recognition database was built from the study of recordings of 30 participants performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors. The objective is to classify activities into one of the six activities performed.
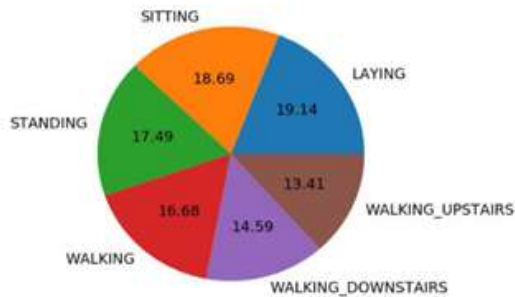
2. Data Preprocessing

Identifying suitable machine learning algorithm to analyze given dataset and analyze the performance of different Machine Learning algorithms and apply these machine learning algorithms to the cleaned dataset. Calculate accuracy rate from the accuracy score.
Compare accuracy rate of different Machine Learning algorithms and Compare time taken by different Machine Learning algorithms.
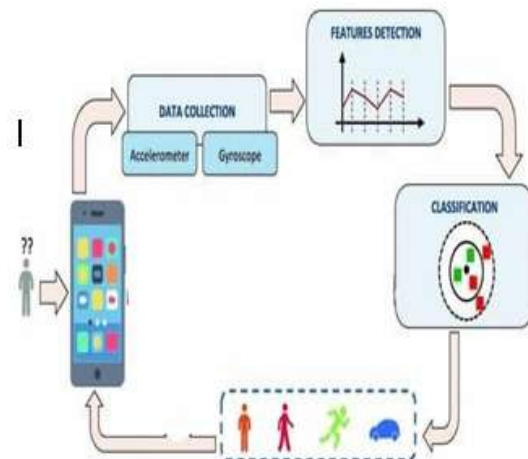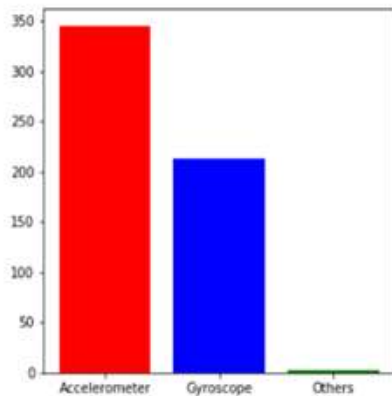

3. Data Visualization

In this we have visualized the training data to get a better understanding of dataset. Firstly we begin by visualizing the percentage of data for each type of activity. Next on visualizing the dataset, we can see there are many features. It's easy to identify that there are accelerometer, gyroscope and other values in the dataset and then we looked for the share of each.

applies. Learn about Machine Learning and its types and approaches.

Learn about various libraries in machine learning and was able to preprocess and visualize a dataset for our model. Choose models according to our requirements and achieve high accuracy, remove outliers in the dataset and link database and monitor and test the model.

4. Classification

In this we have used the various classification models like logistics regression, k-nearest neighbors, random forest, support vector machine and with the help the help of accuracy score we will identify which model will make better predictions with highest accuracy score.
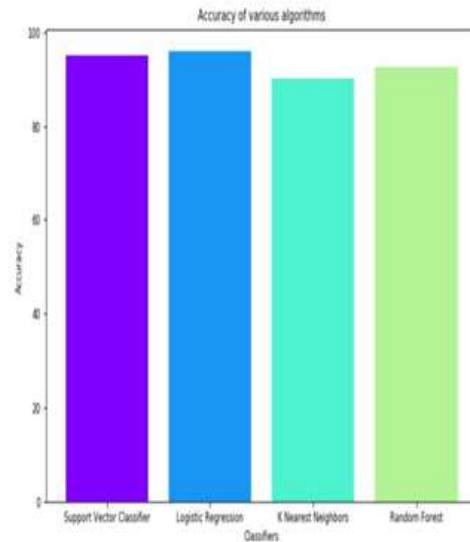
V. RESULT AND DISCUSSION

After object creation and model fitting following are the Predictions we made using accuracy score
ACCURACY OF SVC: 95.045
ACCURACY OF LR: 95.86
ACCURACY OF KNN: 90.023
ACCURACY OF RF: 92.467

VI. CONCLUSION

From this project, we concluded that when we applied numerous machine learning algorithm the logistic regression was the one which classifies the activity with best accuracy as we have seen its accuracy is highest from all the models we have

REFERENCES

[1]. Dataset:Kagglehttps://www.kaggle.com/ucim l/human-activity-recognition-with-smartphones
[2]. Datapreprocessinghttps://towardsdatascience. com/data-preprocessing-concepts-fa946d11c825
[3]. Logisticregressionhttps://www.tutorialspoint. com/machine_learning_with_python/machine learning_with_python_classification_algorith ms_logistic_regression.htm
[4]. RandomForesthttps://www.google.com/url?sa =t&source=web&rct=j&url=https://towardsd atascience.com/understanding-random-forest-
[5]. SVMhttps://www.analyticsvidhya.com/blog/2 017/09/understaing-support-vector-machine-example-code/
[6]. KNNhttps://towardsdatascience.com/machine -learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761