Spring 4-15-2020

Human Activity Recognition & Mobily Path Prediction

Priya Patel

Applied Project in computer Info. Science

Human Activity Recognition & Mobily Path Prediction

Prepared By,

Priya Patel

ID : 225589

Guided By,

Prof. Abrar Qureshi

Harrisburg University of Science and Technology

Table of Content

Table of Figures

Abstract

Individual Mobility is the study that depicts how individuals move inside a region or system. As of late a few researches have been accomplished for this reason and there has been a flood in enormous informational accessible in individual developments. Most of these information's are gathered from cellphone or potentially GPS with variable accuracy relying upon the distance from the tower. Enormous scope information, for example, cell phone follows are significant hotspot for urban modeling. The individual travel designs breakdown into a solitary likelihood distribution however despite the assorted variety of their travel history people follow basic reproducible examples. This similitude in movement example can help us in an extremely different zones of utilizations, for example, city arranging, traffic building, spread of disease and versatile infections. The motive of this project is to show that by utilizing a measure of direct estimation that human directions do follow a few high reproducible scaling designs.

Activity recognition expects to perceive the activities and objectives of at least one operator from a progression of perceptions on the specialists' activities and the natural conditions. Human movement acknowledgment, which is one of the developing fields of research, plans to figure out which action is finished by people. Some true applications, for example, health monitoring, abnormal behavior detection, and sport. In this way, it is a troublesome issue given the enormous number of perceptions delivered each second, the fleeting idea of the perceptions, and the absence of an unmistakable method to relate accelerometer information to known developments. Keen PDAs presently fuse numerous different and ground-breaking sensors, for example, GPS sensors, vision sensors, sound sensors, light sensors, temperature sensors, course sensors and speeding up sensors. This project is about utilizations telephone-based accelerometers to perform activity recognition, which includes identifying the physical movement a user is performing.

Keywords : Mobility, Dataset, GPS, Phone traces, Processing and utilizing data Knowledge Discovery Database, Data Mining,

Chapter 1

Introduction

As of late there has been a high increment being used of cell phones everywhere throughout the world climate it is a cell phone or a GPS tracker. With the assistance of these gadgets a few inquiries about are continuing finding the human development examples or gathering development designs. Now and again, they utilize this information to anticipate the following area of the client. By obtaining these examples, we can understand a colossal number of issues identified with human development, for example, city arranging, traffic the board, crisis reactions, spread of versatile infections and different maladies. With the assistance of cell phone follow information analysts inspect human portability conduct with lower assortment cost, bigger example size as it can have tremendous human base, higher update recurrence and more extensive spatial and worldly inclusion. The cell phone areas are routinely gathered by Google maps and a lot of progressively portable applications as individual travel history; along these lines, the datasets are accessible at no expense. With such many points of interest, the cell phone follow information additionally has a few downsides for investigate. (1) Mobile telephone clients may not speak to an entire populace looked over an arbitrary example. The information ought to be investigated before venturing into an outcome. (2) The datasets are commonly not intended for investigation purposes so for the most part not in a simple to utilize group, which again need some serious handling of the crude information. According to later

Worldwide Attitude Survey there are in excess of 5 billion cell phones over the world and half of them are cell phones. A large portion of the versatile applications running on these PDAs approach consent for area get to, media get to and two or three unique things. Although there is a choice to permit or deny these authorizations, the applications work even more precisely when permitted get to. For instance, on the off chance that we are utilizing a nourishment conveyance application it must know the present area of the individual and gives the nourishment alternatives as needs be. With these procedures running out of sight for the applications, the battery life diminishes in an exponential way. On the off chance that we can kill these procedures when not required will spare the battery life somewhat. This undertaking speaks to the progression towards building a system to use cell phone information or tower area information for discovering the likelihood of the client's area dependent on their recorded way. We will likewise utilize the GPS area information right now check the exactness of anticipated area.

Internet of Things has been viewed as the innovation for flawlessly incorporating old style arranges and organized articles. The essential thought of IoT is to interface everything to the web. The kind of the system which interface anything for example physical articles, gadgets, structures, vehicles and different things installed with programming, sensors and system availability dependent on stipulated conventions that empowers these items to gather and trade information. In our everyday lives, we have gotten increasingly dependent on IoT with our wearable tech, machines, our vehicles.

IoT applications sense the confused condition and creates a tremendous information that must be separated and cleaned so it tends to be deciphered, and client will be given bits of knowledge of the information gathered in type of examples. Across different system foundations, IoT permits detecting of the articles and remotely get to which thusly empowers open doors for a superior combination among genuine and electronic world. mentioned information so it serves the information to the mentioned client with the goal that it can spare part of time. One of the most significant inquiries that emerge now is, how would we convert the information produced or caught by IoT into information to give an increasingly advantageous condition to individuals. These Android telephones, just as essentially all new cell phones and keen music players, including the iPhone and iPod Touch [2], contain tri-pivotal accelerometers that measure quickening in each of the three spatial measurements. These accelerometers are additionally equipped for recognizing the direction of the gadget (helped by the way that they can identify the heading of Earth's gravity), which can give valuable data to action acknowledgment. Accelerometers were at first remembered for these gadgets to help propelled game play and to empower programmed screen pivot, yet they obviously have numerous different applications. There are numerous helpful applications that can be assembled if accelerometers can be utilized to perceive a client's action. We are building a model that precisely groups whether an individual is strolling, strolling upstairs, strolling ground floor or sitting utilizing sensor information.

In recent years there has been a high increase in use of mobile devices all over the world weather it is a mobile phone or a GPS tracker. With the help of these devices several researches are going on finding the human movement patterns or group movement patterns. In some cases, they use these data to predict the next location of the user. By acquiring these patterns, we can solve a huge number of issues related to human movement such as city planning, traffic management, emergency responses, spread of mobile viruses and other diseases. With the help of mobile phone trace data researchers examine human mobility behavior with lower collection cost, larger sample size as it can have huge human base, higher update frequency and broader spatial and temporal coverage. The mobile phone locations are regularly collected by Google maps and many more mobile applications as individual travel history; therefore, the datasets are available at no cost. With so many advantages the mobile phone trace data also have several drawbacks for research. (1) Mobile phone users may not represent a whole population chosen from a random sample. The data should be analyzed before reaching into a result. (2) The datasets are generally not designed for analysis purposes so mostly not in an easy to use format, which again need some intensive processing of the raw data. As per recent

Global Attitude Survey there are more than 5 billion mobile devices across the world and half of them are smartphones. Most of the mobile applications running on these smart phones ask permission for location access, media access and a couple of different things. Though there is an option to allow or deny these permissions, the applications work more accurately when allowed access. For example, if we are using a food delivery application it needs to know the current location of the person and gives the food options accordingly. With these processes running in the background for the applications, the battery life decreases in an exponential way. If we can turn these processes off when not required will save the battery life to some extent.

To address the action acknowledgment task utilizing managed taking in first gathered accelerometer information from four clients as they performed exercises, for example, strolling, rising stairs, dropping stairs, sitting, and standing. At that point totaled this crude time arrangement accelerometer information into models, where every model is marked with the movement that happened while that information was being gathered. At that point manufactured prescient models for movement acknowledgment utilizing grouping calculations

1.1 Problem statement

Several human mobility data have been gathered and published in the past. This mobility data contains rich knowledge about the locations and can help in addressing many challenges. For example, understanding the human mobility behavior inside a city can help forecasting of the traffic. Another example is that we can identify the locations by the means of the transition between these locations, e.g., people usually go to work in the morning and come back home after 4pm on weekdays and visit shopping centers after work or on weekends.

To gather information for action acknowledgment task, it was important to have an enormous information with 4 clients convey an Android-based and iPhone-based PDA while playing out certain ordinary exercises. At that point enrolled the assistance of 4 clients to convey an advanced cell while playing out a arrangement of exercises. They conveyed the PDA in their front leg pocket and were approached to walk, run, rise stairs, slip stairs, sit for explicit timeframes. The information assortment was constrained by an application named Science Journal that was executed on the telephone. This application, which is supported by google. Being a free application, it urges to record perceptions utilizing gadget sensors to quantify light, solid, development and set triggers to advise application when to record. Offers through a basic graphical UI, allowed us to record the triaxial accelerometer's readings, and mark the action being performed. The application allowed us to control what sensor information (e.g., GPS, accelerometer) was gathered and how as often as possible it was gathered. In all cases we gathered the accelerometer information at regular intervals

Chapter 2

Background

In this section I will depict and condense important work that has been done in the field of disseminated picture preparing, at that point portray the model about data collected in mobile devices.

2.1 Relevant work

Paper [1] suggest that A mobile operator can follow individuals' development in urban communities dependent on their cellular network location. This urban human mobility information contains rich information about areas and can help in tending to numerous urban difficulties, for example, traffic blockage or air contamination issues. Right now, review late writing on urban human versatility from an information mining view: from the information assortment and cleaning, to the portability models and the applications. To start with, the author condenses ongoing open urban human portability informational indexes and how to clean and preprocess such information. Second, they depict later urban human versatility models and indicators, e.g., the profound learning indicator, for anticipating urban human portability.

Human Activity Recognition(HAR) is ordering action of an individual utilizing responsive sensors that are influenced from human development. The two clients and capabilities(sensors) of cell phones increment and clients normally convey their cell phone with them. These realities make HAR increasingly significant and famous. This work centers around acknowledgment of human action utilizing cell phone sensors utilizing distinctive AI grouping draws near. Information recovered from advanced cells' accelerometer and spinner sensors are arranged to perceive human movement. Aftereffects of the methodologies utilized are looked at as far as productivity and accuracy.

In this paper[3], author focus on one very important component of aging in place, to be specific movement acknowledgment, which speaks to a framework's capacity to perceive activities performed by clients dependent on a lot of perceptions of their conduct and the earth they end up in (being once in a while additionally alluded to as conduct acknowledgment). It very well may be utilized to follow the conduct of more seasoned grown-ups and guarantee that they carry on in typical parameters. Besides, a smart action acknowledgment framework can likewise recognize when the more seasoned grown-ups are aloof and can prescribe that they move around, go for a stroll, and so on. This should be possible utilizing different faculties like the ones people have. A few arrangements depend on PC vision

In this paper[4], a deep convolutional neural network (convnet) is proposed to perform proficient and viable HAR utilizing cell phone sensors by abusing the innate attributes of exercises and 1D time-arrangement signals, simultaneously giving an approach to consequently and information adaptively remove vigorous highlights from crude information. Trials show that convnets without a doubt determine pertinent and increasingly complex highlights with each extra layer, although distinction of highlight multifaceted nature level declines with each extra layer. A more extensive time length of transient nearby connection can be misused, and a low pooling size is demonstrated to be advantageous.

This exploration gives a comprehensive perspective on human movement acknowledgment framework engineering and examines different issues related with the structure viewpoints. It further endeavors to feature the decrease in computational expense and huge accomplishment in exactness by strategies for highlight determination. It additionally endeavors to present the utilization of repetitive neural systems to take in highlights from the long groupings of time arrangement information, which can contribute towards improving exactness and lessening reliance on area information for include extraction and designing.

Berchtold et al. proposed the Acti Serv stage which utilized a cellphone to catch the quickening signal. The creators built up an effective and convenient fluffy deduction framework to group ambulation exercises. The precision accomplished fluctuated somewhere in the range of 71% and 97%. In the event that the calculation is intended to meet a constant reaction then the precision drops down to 71%, and if the calculation is permitted to prepare to its full limit, which takes a request for days, it arrives at an improved precision of 97%. A subject ward examination supported the exactness to 90%.

There have been a few methodologies proposed to distinguish area following. Most of these methodologies are conventional and may deliver bogus positive, i.e., wrong area distinguished by the framework. Anyway, there has been shockingly constrained work done in the region of understanding the example of individual and gathering portability. Ongoing improvements in area-based advancements empowers us to follow singular development and action investment in urban space across time. Work has likewise been done on frequently visit based area forecast in a quantitative way.

Considering different exact datasets, a few human portability models have been suggested that catches human versatility to a certain degree. There were a significant couple of explores where they utilized the cellphone information as GPS was not accessible on the cell phones in mid twentieth century. Thus, analysts utilized the accessible pinnacle information source. Because of the critical utilization of cell phones, the investigation of human portability has changed a great deal. Current cell phones use cell tower data and the GPS framework for increasingly exact area following. With billions of individuals conveying their telephones ordinary gives a lot of information on human development. This information is being gathered and made accessible as open celled or Mozilla area administration, open new open doors for displaying and anticipating human portability more precisely.

There were quite a couple of researches where they used the cellphone data as GPS was not available on the mobile phones in early 20th century. So, researchers used the available tower data source. Due to the significant use of mobile phones, the study of human mobility has changed a lot. Current mobile phones utilize cell tower information and the GPS system for more accurate location tracking. With billions of people carrying their phones everyday provides a large amount of data on human movement. This data is being collected and made available in the form of open cellid or Mozilla location service, open new opportunities for modelling and predicting human mobility more accurately.

Sensors in IoT applications sense the confused condition and creates a tremendous information that must be separated and cleaned so it tends to be deciphered, and client will be furnished with bits of knowledge of the information gathered in type of examples. Across different system frameworks, IoT permits detecting of the items and remotely get to which thus empowers open doors for a superior incorporation among genuine and modernized world. mentioned information so it serves the information to the mentioned client with the goal that it can spare part of time. One of the most significant inquiries that emerge now is, how would we convert the information produced or caught by IoT into information to give a progressively advantageous condition to individuals. These Android telephones, just as basically all new cell phones and brilliant music players, including the iPhone and iPod Touch

There have been a few methodologies proposed to identify area following. The greater part of these methodologies is conventional and may create bogus positive, i.e., wrong area identified by the framework. Anyway, there has been shockingly restricted work done in the territory of understanding the example of individual and gathering versatility. Ongoing advancements in area-based advances empowers us to follow singular development and movement cooperation in urban space across time. Work has likewise been done on frequently visit based area expectation in a quantitative way.

Considering different exact datasets, a few human versatility models have been suggested that catches human portability to a certain degree. There were a significant couple of explores where they utilized the cell phone information as GPS was not accessible on the cell phones in mid twentieth century. In this way, analysts utilized the accessible pinnacle information source. Because of the huge utilization of cell phones, the investigation of human versatility has changed a ton. Current cell phones use cell tower data and the GPS framework for increasingly precise area following. With billions of individuals conveying their telephones ordinary gives a lot of information on human development. This information is being gathered and made accessible as open cell-id or Mozilla area administration, open up new open doors for displaying and anticipating human portability all the more precisely.

In literature [8] A mobile operator can follow individuals' development in urban communities dependent on their cellular network location. This urban human mobility information contains rich information about areas and can help in tending to numerous urban difficulties, for example, traffic blockage or air contamination issues. Right now, review late writing on urban human versatility from an information mining view: from the information assortment and cleaning, to the portability models and the applications. To start with, the author condenses ongoing open urban human portability informational indexes and how to clean and pre-process such information. Second, they depict later urban human versatility models and indicators, e.g., the profound learning indicator, for anticipating urban human portability.

This paper [9] shows Most of these methodologies are conventional and may deliver bogus positive, i.e., wrong area distinguished by the framework. Anyway, there has been shockingly constrained work done in the region of understanding the example of individual and gathering portability. Ongoing improvements in area-based advancements empowers us to follow singular development and action investment in urban space across time. Work has likewise been done on frequently visit based area forecast in a quantitative way.

There have been several approaches proposed to detect location tracking. Most of these approaches are generic and may produce false positive, i.e., wrong location detected by the system. However there has been surprisingly limited work done in the area of understanding the pattern of individual and group mobility. Recent developments in location-based technologies enables us to track individual movement and activity participation in urban space across time. Work has also been done on frequently visit based location prediction in a quantitative manner. Based on various empirical datasets, several human mobility models have been proposed that captures human mobility to a certain extent.

Sensors in IoT applications sense the complicated environment and generates an enormous data that must be filtered and cleaned so that it can be interpreted, and user will be provided with insights of the data collected in form of patterns. Across various network infrastructures, IoT allows sensing of the objects and remotely access which in turn enables opportunities for a better integration between real and computerized world. requested data so it serves the data to the requested user so that it can save lot of time. One of the most important questions that arise now is, how do we convert the data generated or captured by IoT into knowledge to provide a more convenient environment to people. All these Android phones, as well as virtually all new smartphones and smart music players, including the iPhone and iPod Touch [2], contain tri-axial accelerometers that measure acceleration in all three spatial dimensions. These accelerometers are also capable of detecting the orientation of the device (helped by the fact that they can detect the direction of Earth’s gravity), which can provide useful information for activity recognition. Accelerometers were initially included in these devices to support advanced game play and to enable automatic screen rotation, but they clearly have many other applications. In fact, there are many useful applications that can be built if accelerometers can be used to recognize a user’s activity. We are building a model that accurately classifies whether an individual is walking, walking upstairs, walking downstairs or sitting using sensor data.

In order to address the activity recognition task using supervised learning first collected accelerometer data from four users as they performed activities such as walking, ascending stairs, descending stairs, sitting, and standing. Then aggregated this raw time series accelerometer data into examples, where each example is labeled with the activity that occurred while that data was being collected. Then built predictive models for activity recognition using classification algorithms.

2.2 Mobility Dataset

Several human mobility data have been gathered and published in the past. This mobility data contains rich knowledge about the locations and can help in addressing many challenges. For example, understanding the human mobility behavior inside a city can help forecasting of the traffic. Another example is that we can identify the locations by the means of the transition between these locations, e.g., people usually go to work in the morning and come back home after 4pm on weekdays and visit shopping centers after work or on weekends.

The main mobility datasets are recorded according to 1. relevant location with access points (Cellular tower, Wifi access points etc.), 2. GPS information by devices, 3. Aggregated GPS points recorded by vehicles such as taxis or buses. For this project work we collected the mobile data for a couple of months using several mobile applications such as Tower Collector, Network Cell info

Lite, GPS logger etc. We will cover them one by one below.

Tower Collector: This mobile application can run in background and collect the data. The collected data consists of measurements about the Tower, Network type, cell id, latitude and longitude of the tower etc. Once the data is collected, we can retrieve it using different file formats like CSV, GPX or JSON.

The advantage about the Tower Collector application is that it collects all the required information and stores them cumulatively without losing and old data from the data sets

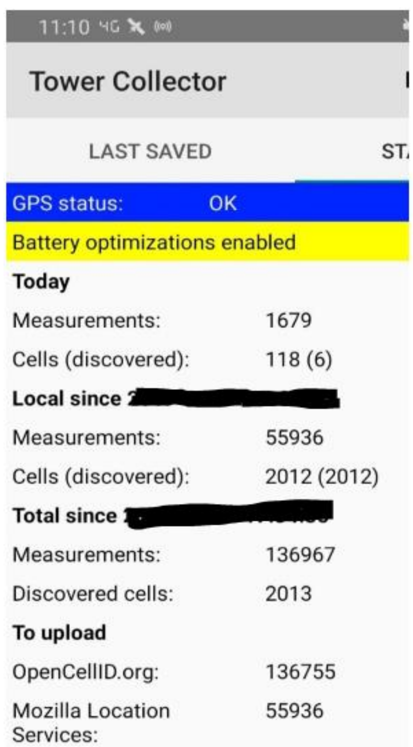


Fig 1. Tower collector data application

Application collects the data of individuals movement like walking, running, sitting

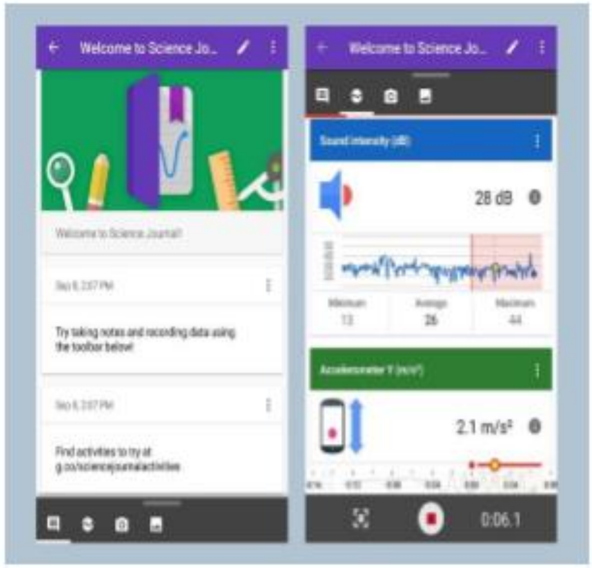


Fig2. Application which collect the physical movement

GPS Logger: GPS logger is a mobile application that records the geo-coordinates of the mobile. We have the option to get the data as per our need by setting the time interval to 1 sec or 5 sec or something else. This data includes the attributes like user's current location coordinates (latitude and longitude), date, time etc.

The main purpose of using this application is to the GPS coordinates at the specified intervals to a file and upload it to cloud automatically. The logs can be saved in the formats like GPX, KML, CSV or NMEA files.

It is important to compare the performance of multiple different machine learning algorithms consistently. When you have a new dataset, it is a good idea to visualize the data using different techniques in order to look at the data from different perspectives. the same idea applies to model selection. You should use a number of different ways of looking at the estimated accuracy of your machine learning algorithms in order to choose the one or two to finalize.

A way to do this is to use different visualization methods to show the average accuracy, variance and other properties of the distribution of model accuracies.The key to a fair comparison of machine learning algorithms is ensuring that each algorithm is evaluated in the same way on the same data.

Logistic Regression

Linear Discriminant Analysis

K-Nearest Neighbors

Classification and Regression Trees

Naive Bayes

Support Vector Machines

The problem is a standard binary classification dataset called the Pima Indians onset of diabetes problem. The problem has two classes and eight numeric input variables of varying scales.

Data Collection:

In order to collect data for activity recognition task, it was necessary to have a large data with 4 users carry an Android-based and iPhone based smart phone while performing certain everyday activities. Then enlisted the help of 4 users to carry a smart phone while performing a specific set of activities. They carried the cell phone in their front leg pocket and were asked to walk, jog, ascend stairs, descend stairs, sit for specific periods of time. The data collection was controlled by an application named Science Journal that was executed on the phone. This application, which is backed by google. Being a free application, it encourages to record observations using device sensors to measure light, sound, movement and set triggers to tell app when to record. Offers through a simple graphical user interface, permitted us to record the triaxial accelerometer's readings, and label the activity being performed. The application permitted us to control what sensor data (e.g., GPS, accelerometer) was collected and how frequently it was collected. In all cases we collected the accelerometer data every 5 seconds

Measuring Acceleration in the Science Journal App

Your phone has a built-in device called an accelerometer that measures acceleration, which is the change in an object's velocity per second. Since velocity is measured in meters per second (m/s), acceleration is measured in meters per second per second, which is written as meters per second squared (m/s2). Accelerometers are used for things like motion controls in video games or to detect when you pick your phone up.

Unlike the light sensor and microphone, which each just record one value, the accelerometer on your device records acceleration in three different directions, or "axes" (orientations in space), referred to as X, Y, and Z. Even though the accelerometer is a single sensor built into your device, in the Science Journal app, there is a separate accelerometer sensor for each axis.

If you look in the Science Journal app along the strip of sensors, you will find Accelerometer X, Accelerometer Y, and Accelerometer Z. (There is also an accelerometer sensor called the Linear accelerometer, which will be talked about later in this tutorial.) The X, Y, and Z axes correspond to a physical direction relative to your phone's body, as shown in the diagram below.[11]

2.2.1 Ease of Use

Path mobility solve a huge number of issues related to human movement such as city planning, traffic management, emergency responses, spread of mobile viruses and other diseases. Human activity recognition, which is one of the growing fields of research, aims to determine which activity is done by individuals. Many real-world applications such as health monitoring, abnormal behavior detection, and sport. Across various network infrastructures, IoT allows sensing of the objects and remotely access which in turn enables opportunities for a better integration between real and computerized world. requested data so it serves the data to the requested user so that it can save lot of time and cost.

Activity recognition expects to perceive the activities and objectives of at least one operator from a series of perceptions on the specialists' activities and the natural conditions. Human action acknowledgment, which is one of the developing fields of research, plans to figure out which movement is finished by people.

Singular mobility is the examination that portrays how individuals move inside a zone or system. As of late a few inquiries about have been accomplished for this reason and there has been a flood in huge dataset available in singular developments. A large portion of these datasets are gathered from cellphone as well as GPS with variable precision relying upon the good ways from the tower. Huge scope information, for example, cell phone follows are significant hotspot for urban modeling.

Here are some steps to describe the design of the system step by step.

Step 1: we collect a data of individuals movement from using several mobile applications such as Tower Collector, Network Cell info Lite, GPS logger, etc.

Step 2: This application will capture the GPS coordinates at the specified intervals to a file and upload it to cloud automatically. The logs can be saved in the formats like GPX, KML, CSV or NMEA files.

Step 3: After collecting the data of the tower location & mobile location MYSQL is used to merge them at a single dataset by means of matching the date and time

Step 4: This dataset includes the fields like Cell_id, Latitude, Longitude, Glongitude, Glatitude, Glatlon, date and time. Glatlon is a created ID that is used to represent a single GPS Latitude and longitude.

Another goal of the prediction of system is prediction of human activity Recognition.

Step 1: collect the data of human behavior using mobile application Science Journal.

Step 2:this application provides the backup by Google and store the human movement data.

Step 3: it records the observation with different sensors like light, sound, movement and set the triggers to set the application on record.

Step 4: The data collected has five columns. 3 columns for the x, y, z axis and user column , the activity column.

Step5: Collected data was divided into 2 categories: one for training of system & second for evaluating the performance of system

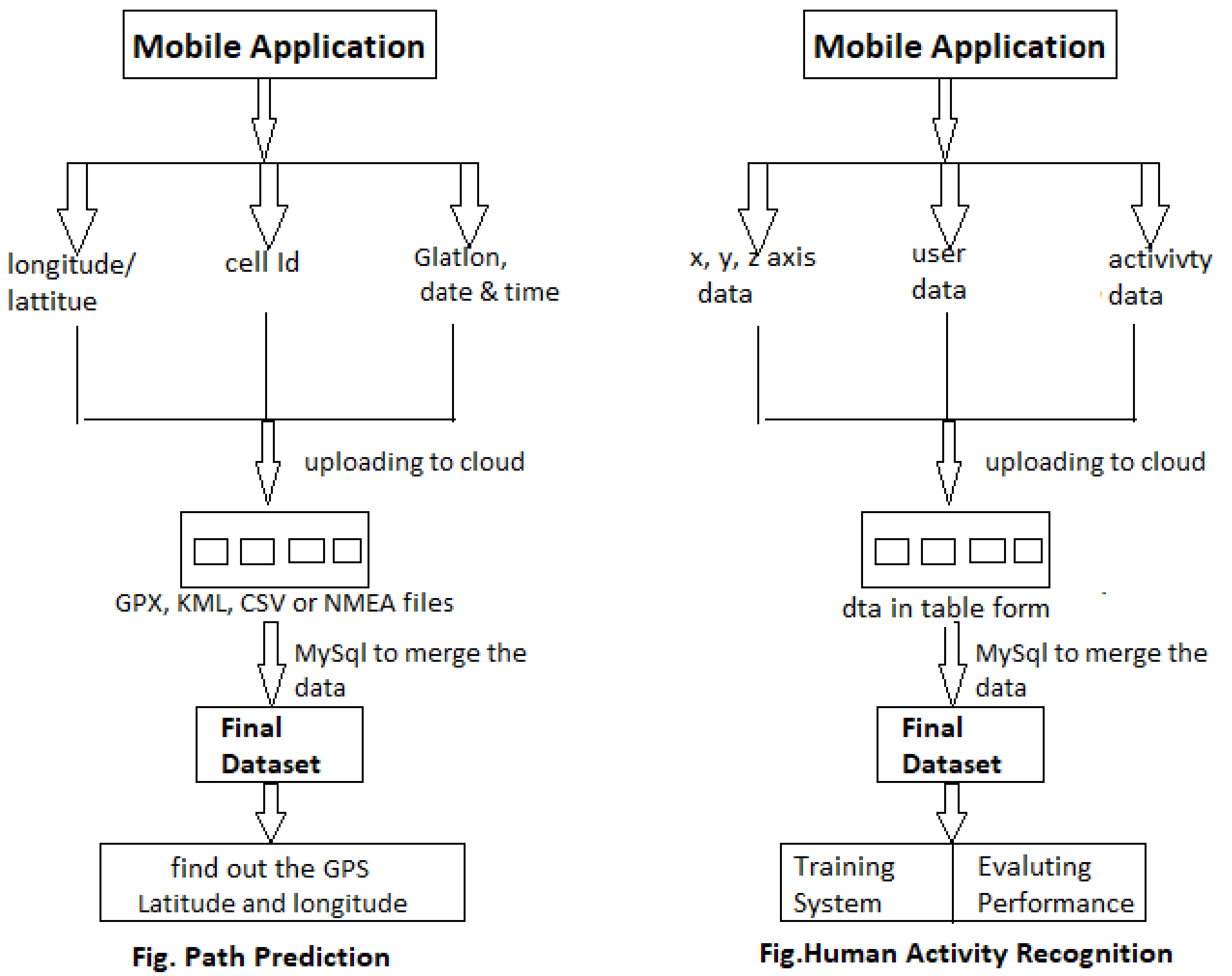


Fig 3. Detailed Design of the System

2.3 Mobility Data Preprocessing

After collecting both the data about the tower location and mobile location we used MYSQL to merge them at a single dataset by

matching the date and time. This is the final data set used for all the future work. This dataset includes the fields like Cell_id, GLatLon, Latitude, Longitude, Glatitude, Glongitude, date and time. Glatlon is a created ID that is used to represent a unique GPS Latitude and longitude.

mcc	mnc	lac	cell_id	psc	asu	dbm	ta	lat	lon	accuracy	speed	bearing	altitude
	4162	5	504					42.66205	-73.7733	8.58	0	0	45.89
310	120	5633	46797361	176	25	-115		42.66205	-73.7733	8.58	0	0	45.89
	4162	5	436					42.662	-73.7733	10.72	0	0	50.87
310	120	5633	46797361	176	23	-117		42.662	-73.7733	10.72	0	0	50.87
	4162	5	504					42.66199	-73.7733	6.43	0	0	43.37
310	120	5633	46797361	176	20	-120		42.66199	-73.7733	6.43	0	0	43.37
	4162	5	504					42.66199	-73.7733	7.5	0	0	51.31
310	120	5633	46160641	51	43	-97		42.66199	-73.7733	7.5	0	0	51.31
	4162	5	504					42.66199	-73.7733	5.36	0	0	41.75
310	120	5633	46797361	176	25	-115		42.66199	-73.7733	5.36	0	0	41.75
	4162	5	436					42.66199	-73.7733	8.58	0	0	53.99

Fig 4.Processed data retrieved from MYSQL

After collecting both the data about the tower location and mobile location we used MYSQL to merge them at a single dataset by matching the date and time. This is the final data set used for all the future work.

Chapter 3

Data Exploration

The data collected has 5 columns initially. Three columns for the x, y, z axis then user column and the activity column. There are 4 activities Sitting, Walking, Upstairs, Downstairs which we are going to recognize. Both android and iPhone device is used for this project. When the activity is stationery like sitting, laying, standing then android tends to record few accelerometer readings as missing value. Whereas iPhone gives no missing value for any accelerometer. So to fix the missing value within the dataset all missing value are filled with zero

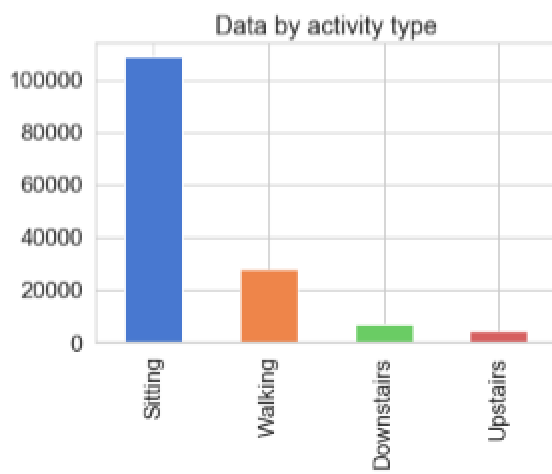


Fig 5: Data Representation of physical movement

3.1 Software Requirement

I. Mobile devices
II. Mobile application science journal (to collect human activity)
III. Applications such as tower collector, network cell info lite, GPS logger, etc. (To collect data like location longitude, latitude)
IV. MYSQL software

3.2 Implementation

The z-axis captures the forward movement of the leg, y-axis captures the upward and downward motion and the x-axis captures the horizontal movement of user's leg. The acceleration of the activities can be easily understood from time series graph of each activity.

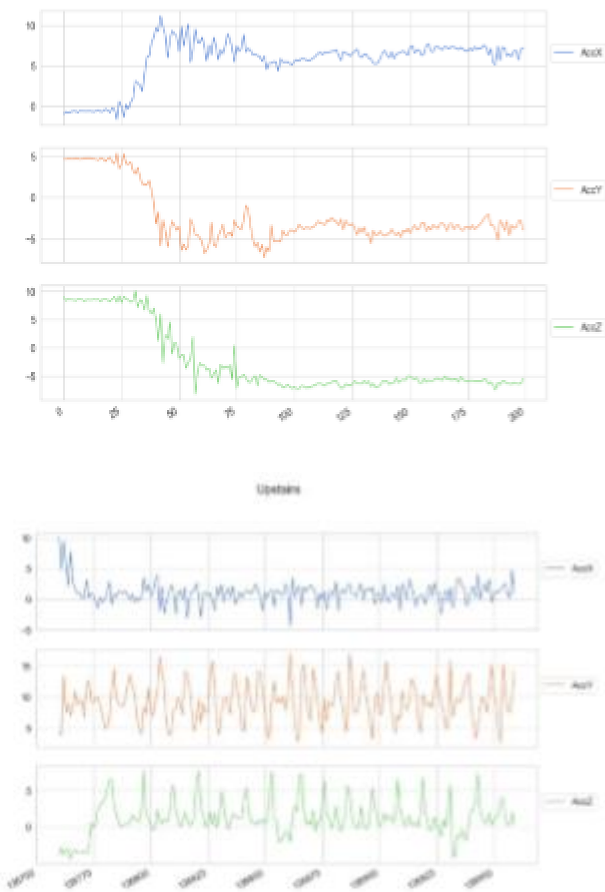


Fig 6 Graph Representation of physical movement

Ideally data should be something like this. the activity recognition algorithm should be capable of classifying the data corresponding to each activity. The task of activity recognition is considered as a classification problem where each activity corresponds to one class. The features extracted is then classified into different activities.

Chapter 4

Execution Procedure:

Prediction: In this study, next location prediction is based on a history of visits for a user. We used random forest model to predict the next location. Random forest consists of a large no of individual decision trees that operate as an ensemble. Each prediction tree in this model spits out a class prediction and the class with the most votes becomes our model's prediction.

To test the model first we used 70% of the data to train the model and the remaining 30% for testing. We tested it by predicting the GPS location based on the Tower location.

The accuracy in this case was near to 44% (Fig 7) which was a setback to this model. After careful review we found that visits at different locations are highly influenced by the day of the week, the time of the day. So, we added the time attribute as a feature to the model. This can be justified as a person can have different paths for a large range of days, but the location can be dependent on the time. For example, if we are taking about an employee then he will be at office between 8am to 4.30 pm on weekdays and will be at home after 7 pm mostly. So, by taking the time as a feature we got an accuracy near to higher 80% (Fig 8).

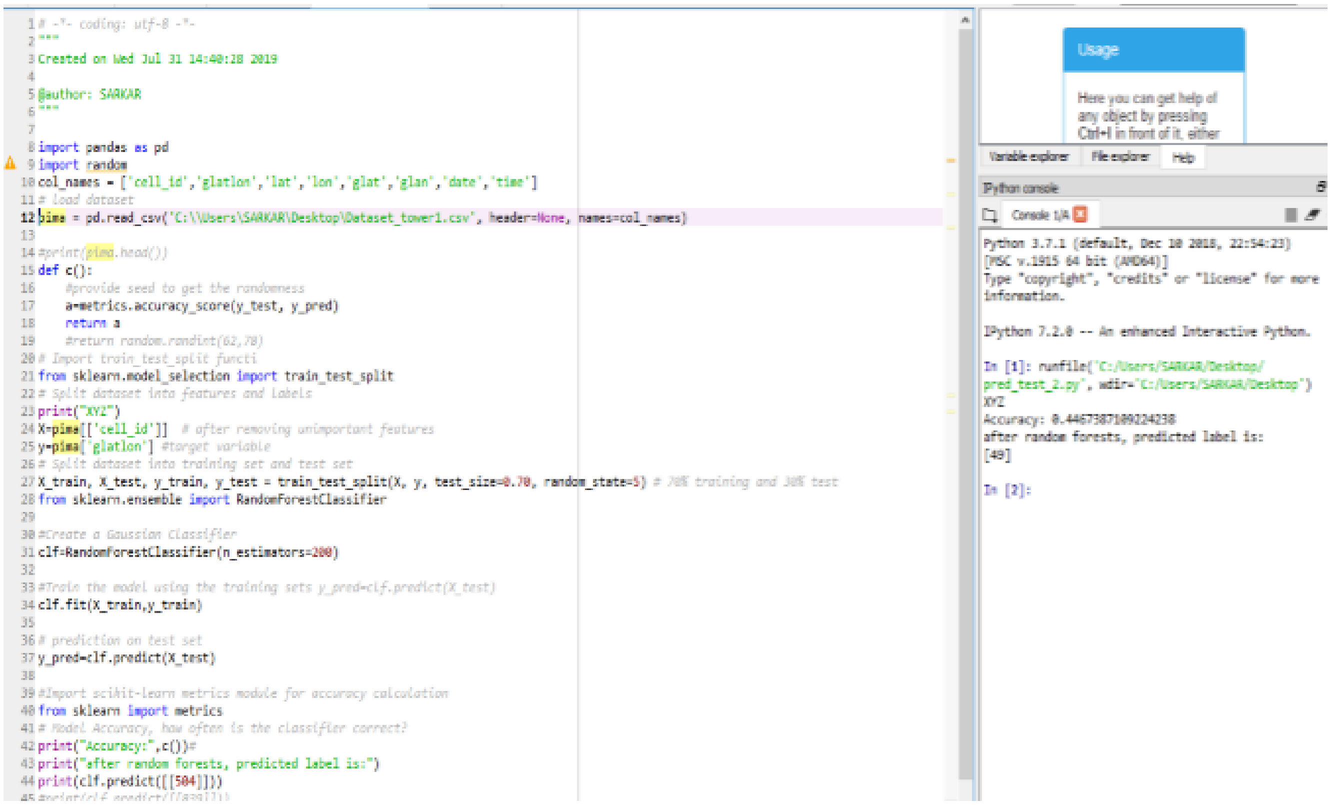


Fig 7Prediction of next location based on only tower information

We tried to plot the geo-coordinates on a graph and found the below results, which depicts the relationship between the Tower location and GPS location.

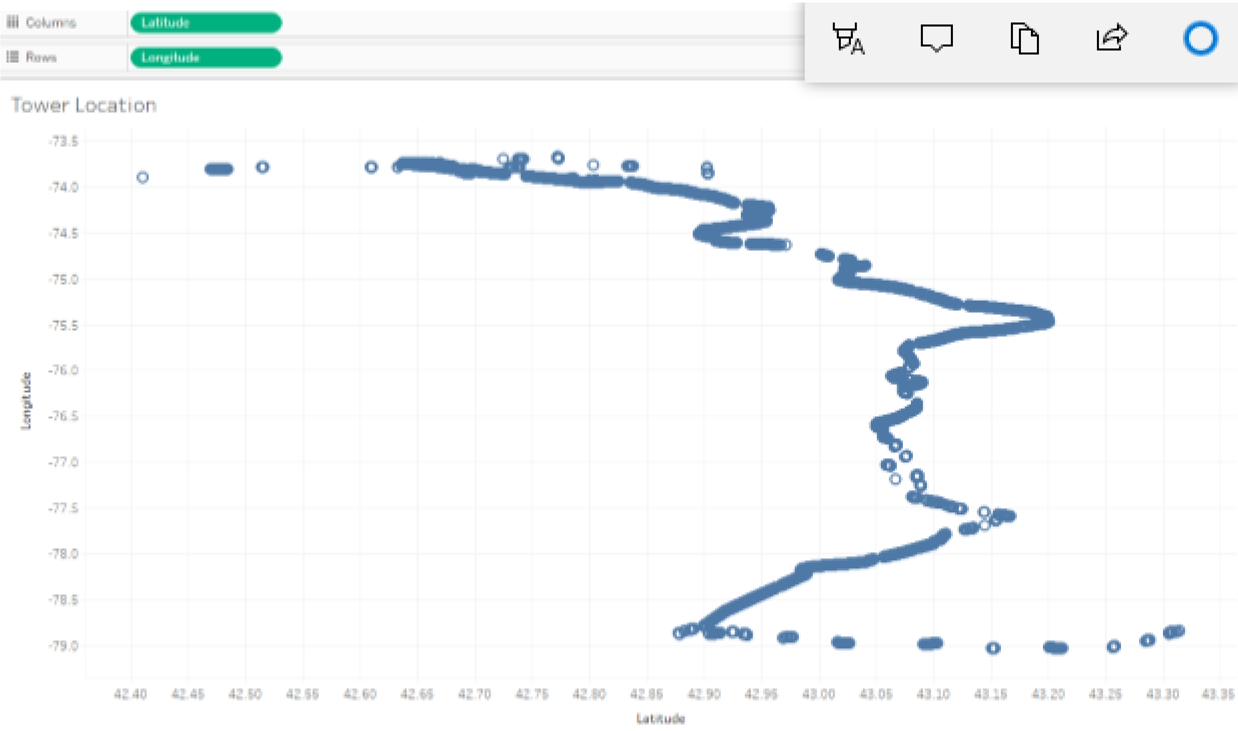


Fig 8: Coordinates of the visited tower locations

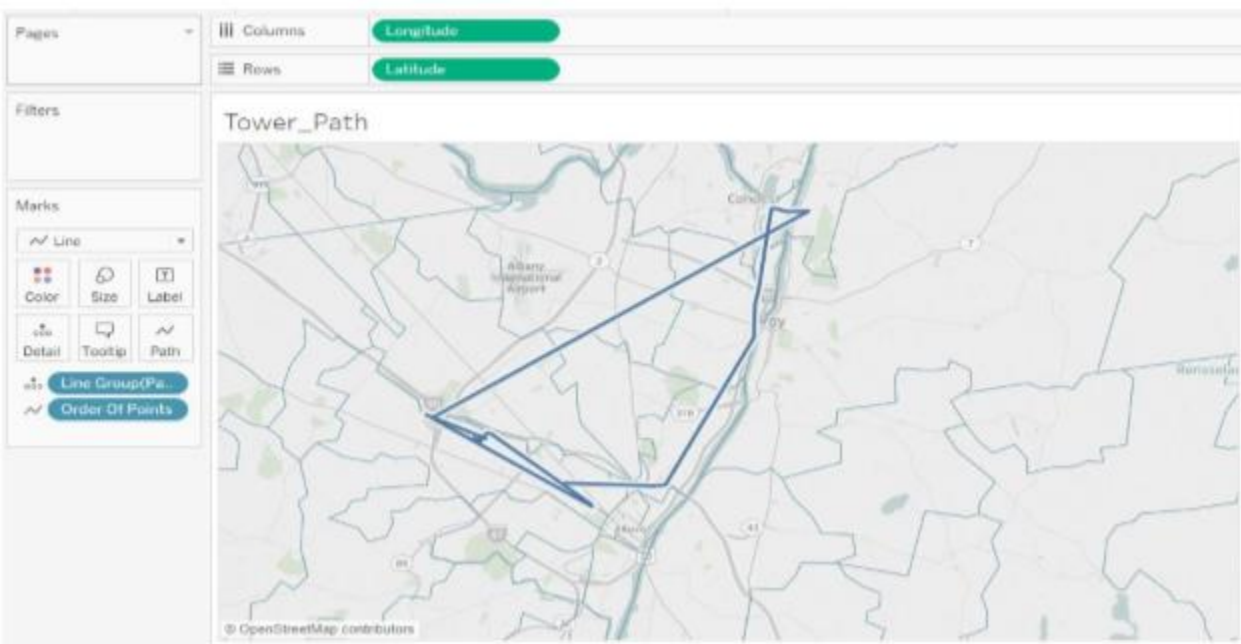


Fig 9 : Connected travel path based on Tower location

We also tried to plot the geo-coordinates of the Tower locations and GPS locations on the map for a specific day and the results are shown in Fig 11 and Fig 12. This also brings us to the conclusion that we can get the human mobility location with the help of cell tower data.

Experiment

We added two more columns UserEncoded and ActivityEncoded. With corresponding encoded value for the column value of activity and username. This is needed because the model which we will build cannot work with non-numerical labels.

The collected data was divided into two categories: one for training of the system and the other for evaluating its performance(testing). It is important to separate the whole data set into a training set and a test set. However, when we decide to split the data, we never want information from the test set to bleed into your training set. This might be great for the overall performance of the model during training and then validation against the test set. But our model is very unlikely to generalize well for data it has not seen yet. The idea behind splitting is we want our model to learn from training set which have been through the experiment. Next, we then want to see how well our model predicts the movements of persons it has not seen before. In our experiment, we used 70% of the data for training and rest 30% for testing. The selection of training and testing sets were based on selecting features using the training variable and the target variable. The training set being the triaxial accelerometers along with username and testing set the activity.

Model used

We apply different algorithms on the train dataset and evaluate the performance on the test data to make sure the model is stable. Constructing highly accurate classifier that generalizes well on data. Several models were used and based on their accuracy and classification report we obtained results.

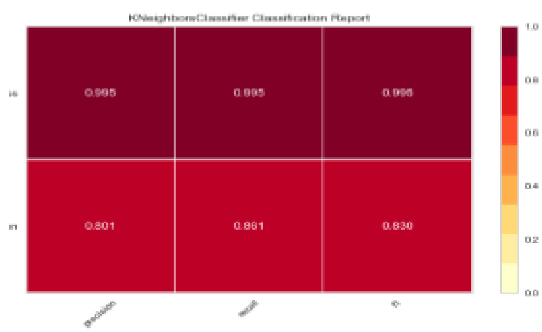
Naïve Bayes Algorithm Accuracy: 65.5%

Linear SVC Accuracy: 83%

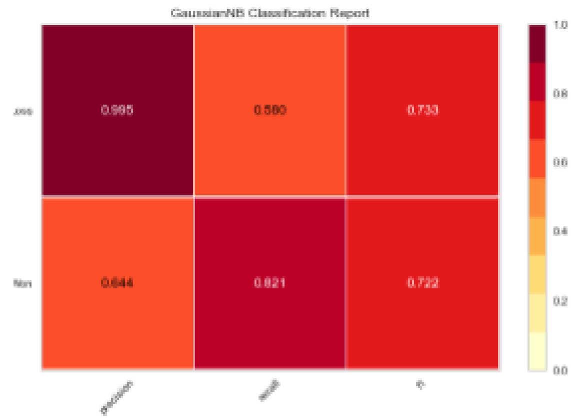
K Neighbors Accuracy: 97%

```python
# 3 K Neighbours Classifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score
#create object of the lassifier
neigh = KNeighborsClassifier(n_neighbors=3)
#Train the algorithm
neigh.fit(data_train, target_train)
# predict the response
pred = neigh.predict(data_test)
# evaluate accuracy
print ("KNeighbors accuracy score : ",(accuracy_score(target_test, pred))*100)
```

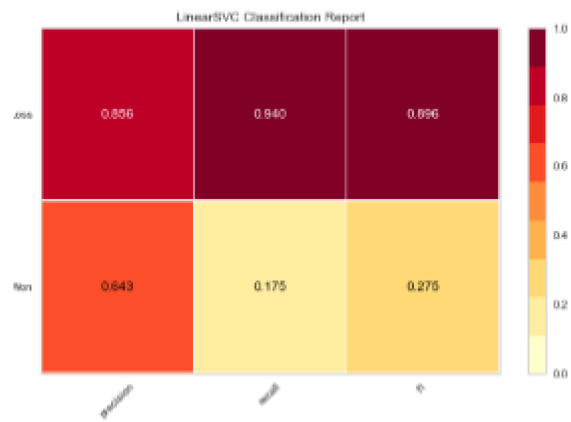
```python
# 1. Naive Bayes
from yellowbrick.classifier import ClassificationReport
# Instantiate the classification model and visualizer
visualizer = ClassificationReport(gnb, classes=['Won','Loss'])
visualizer.fit(data_train, target_train) # Fit the training data to the visualizer
visualizer.score(data_test, target_test) # Evaluate the model on the test data
g = visualizer.poof() # Draw/show/poof the data
```



K- Neighbors Classification Report



Naïve Bayes Classification Report



Linear SVC Classification Report

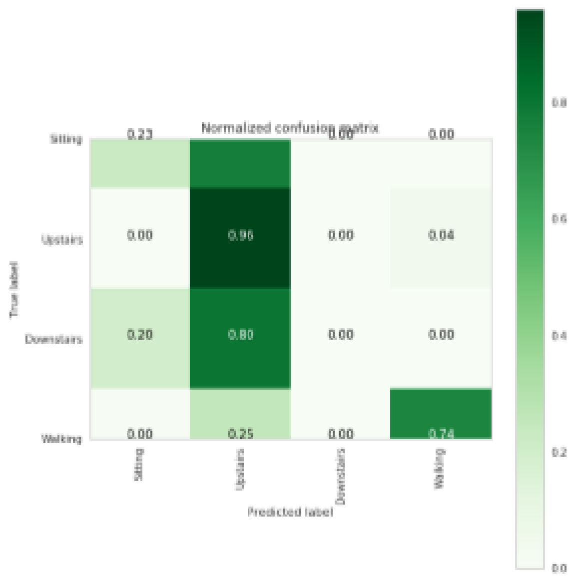
Some More Algorithms

Logistic Regression

```
---------------------
|      Accuracy      |
---------------------

    0.855212572225095

--------------------
| Confusion Matrix |
--------------------
```



Decision Tree

```
---------------------
|      Accuracy      |
---------------------

    0.9713347872029497


--------------------
| Confusion Matrix |
--------------------
```


```
-------------------------
| Classifiction Report |
-------------------------
              precision    recall
f1-score   support

           0       0.62      0.23
0.34      2175
           1       0.87      0.96
0.91     32614
           2       0.00      0.00
0.00      1313
           3       0.83      0.74
0.78      8377

    accuracy
0.86     44479
   macro avg       0.58      0.48
0.51     44479
weighted avg       0.82      0.86
0.83     44479
```

Confusion Matrix in Machine Learning

In the field of machine learning and specifically the problem of statistical classification, a confusion matrix, also known as an error matrix. A confusion matrix is a table that is often used to describe the performance of a classification model (or “classifier”) on a set of test data for which the true values are known. It allows the visualization of the performance of an algorithm.

It allows easy identification of confusion between classes e.g. one class is commonly mislabeled as the other. Most performance measures are computed from the confusion matrix.

A confusion matrix is a summary of prediction results on a classification problem. The number of correct and incorrect predictions are summarized with count values and broken down by each class. This is the key to the confusion matrix. The confusion matrix shows the ways in which your classification model is confused when it makes predictions. It gives us insight not only into the errors being made by a classifier but more importantly the types of errors that are being made.

	Class 1 Predicted	*Class 2 Predicted*
Class 1 Actual	**TP**	**FN**
Class 2 Actual	**FP**	**TN**

High recall, low precision: This means that most of the positive examples are correctly recognized (low FN) but there are a lot of false positives.

Low recall, high precision: This shows that we miss a lot of positive examples (high FN) but those we predict as positive are indeed positive (low FP)

F-measure:

Since we have two measures (Precision and Recall) it helps to have a measurement that represents both of them. We calculate an F-measure which uses Harmonic Mean in place of Arithmetic Mean as it punishes the extreme values more.

The F-Measure will always be nearer to the smaller value of Precision or Recall.

Chapter 5

Conclusion :

Today, 50% of the world's population lives in cities and it will raise to 70% by 2050. Understanding the human mobility is crucial for urban planning, traffic forecasting, epidemic control and many more. It is estimated that more than 5 billion people have mobile phones and over half of these connections are smartphones. With the help of GPS in these mobile devices we can locate the device or person accurately, but we can also locate them with the help of Network data or cellular connectivity. This will certainly decrease the cost in some extent. Though GPS does not use data, but navigation applications that require a server connection do use data.

A portion of the human day by day exercises were sorted utilizing the accelerometer information in an advanced mobile phone. Right now, portrayed how an advanced mobile phone can be utilized to perform action acknowledgment, basically by keeping it in pocket. Furthermore, these exercises can be perceived rapidly, since every model is created worth of information.

We can improve our action acknowledgment in a few different ways. The direct enhancements include: 1) figuring out how to perceive extra exercises, for example, bicycling and vehicle riding, 2) getting preparing information from more clients with the desire that this will improve our outcomes, 3) producing extra and increasingly refined highlights while amassing the crude time-arrangement information, and 4)evaluating the effect of conveying the wireless in various areas

References

[1] Kai zhao, sasu tarkoma, siyuan liu and huy vo, urban human mobility data mining: an overview. IEEE international conference on big data (dec 2016)

[2] Erhan bulbul, aydin cetin, ibrahim alper dogru, human activity recognition using smartphones, 2018 2nd international symposium on multidisciplinary studies and innovative technologies (ismsit)

[3]. G. Tanaka, m. Okada, and h. Minemo, "gps-based daily context recognition for lifelog generation using smartphone," in (ijacsa) int. Journal of advanced computer science and applications

[4] Mo L., Li F., Zhu Y., Huang A. Human physical activity recognition based on computer vision with deep learning model; Proceedings of the 2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings (I2MTC)

[5] Panel Charissa Ann, Ronao Sung-Bae Cho, Human activity recognition with smartphone sensors using deep learning neural networks published in science direct Volume 59, 15 October 2016, Pages 235-244

[6] Mustafa Badshah, Sensor - Based Human Activity Recognition Using Smartphones Spring 5-20-2019

[7] Augustin Chaintreau, Pan Hui, Jon Crowcroft, Fellow, IEEE, Christophe Diot, Richard Gass, and James Scott, Impact of Human Mobility on Opportunistic Forwarding Algorithms. IEEE Transactions on Mobile Computing (June 2007)

[8] Faina Khoroshevsky and Boaz Lerner, Human Mobility-Pattern Discovery and Next-Place Prediction from GPS Data. MPRSS 2016

[9] Learning Model Building in Scikit-learn: A Python Machine Learning Library Geeks for Geeks

[10] https://www.geeksforgeeks.org/confusion-matrix-machine-learning/

[11] Jason Bornley, A Gentle Introduction to Scikit-Learn: A Python Machine Learning Library April 2014.